

THIS WEEK

EDITORIALS

SHOOTINGS Biologist Amy Bishop escapes death penalty for murders **p.6**

WORLD VIEW Gene-therapy success demands a new regulatory approach **p.7**



VAMPIRE SQUID Plankton and dead crustaceans, but hold the garlic **p.8**

The price of progress

A new uranium enrichment technique approved by the US Nuclear Regulatory Commission could have an impact on nuclear proliferation. This should have been taken into account.

Scientists and scientific journals are naturally inclined towards progress. But there are exceptions: progress is not always better. Last week, the US Nuclear Regulatory Commission (NRC) in Rockville, Maryland, announced that it had issued a licence for a new technology to enrich uranium using lasers — a process that is potentially cheaper and better than current enrichment methods, but also politically fraught and possibly dangerous (see *Nature* <http://doi.org/jfr; 2012>). The decision was unfortunate. The NRC should introduce rules to ensure that its future moves are better informed.

Nuclear power and nuclear weapons both require uranium that has been enriched to contain higher levels of uranium-235 than occurs naturally. The 1970 nuclear non-proliferation treaty gives countries the right to possess enrichment systems, but most are subject to strict controls. Where facilities are declared as peaceful, they are subject to inspection by the International Atomic Energy Agency, which ensures that material is not diverted for military or criminal purposes. National intelligence agencies, too, keep a careful eye on the activities of friends and rivals worldwide.

Current techniques to separate uranium-235 from its heavier, non-fissile sibling uranium-238 are hard to hide. They typically involve enormous gaseous-diffusion plants, or smaller, energy-intensive centrifuge facilities that spin the two isotopes apart in high-speed machines. Both require lots of space and electricity, which means that such facilities are easy to spot with, say, satellite reconnaissance.

Lasers are more discreet. When finely tuned, they can ionize uranium-235 that can then be collected on a negatively charged plate. The new technique will require considerably less space and electricity than centrifuge or diffusion plants.

GE Hitachi, the multinational company pursuing laser enrichment, describes it as a “game-changing technology”, and those who fear nuclear proliferation agree. Understanding where enrichment facilities are and how they operate is crucial to maintaining the treaties and agreements that limit the spread of nuclear technology. When undeclared enrichment facilities are uncovered, as happened in Iran in 2009, it can spark an international crisis. Even more worrying is the possibility that some facilities can simply never be detected. Mistrust could lead to regional arms races and even open conflicts.

When the NRC approved GE Hitachi’s application for a laser enrichment plant, it considered physical security, such as the height of the fence around the plant. But the commission did not explicitly address the existential threat that a plant might pose to the non-proliferation regime, nor did it concern itself with the message that the pursuit of this technology might send to other nuclear nations.

The American Physical Society (APS), based in College Park, Maryland, would like to change that. Backed by environment groups and scientists, the society has proposed that the NRC change its regulations to require future applicants to submit a ‘proliferation review’. This would force the commission and corporations to consider the

proliferation consequences of their proposed facilities. It would also provide a public focus for debate, much as environmental-impact statements have allowed activists and citizens to query the environmental aspects of construction.

Could such a review have changed the outcome of the NRC’s licence decision? Perhaps. From a commercial standpoint, the benefit of laser enrichment is clear, but from a broader societal perspective, it is questionable. The cost of fuel for nuclear reactors is dwarfed by the cost

“To consider proliferation might seem beyond the remit of a regulator, but it is not.”

of construction, and this has stalled a nuclear renaissance in the West. A 2010 analysis showed that the savings to households would be less than US\$10 per year from laser enrichment — a saving that many homeowners might be unwilling to accept if presented with the risks the technology poses (see F. Slakey and L. R. Cohen *Nature* **464**, 32–33; 2010).

The APS’s proposed rule change will be voted on by the five NRC commissioners in November. *Nature* urges the commissioners (who were not directly involved in approving the laser enrichment system) to approve the change. To consider something as vague as proliferation might seem beyond the remit of a regulator, but it is not. Under the Atomic Energy Act of 1954, the NRC is charged with considering whether its licences would “be inimical to the common defense and security” of the US public. Nuclear technologies that are potentially destabilizing pose a threat to the security of citizens everywhere, and so the commission could rightly consider it under an altered licensing process. In the case of laser enrichment, as with other technologies, progress should be judged on more than the technical merits alone. ■

Power cuts

China’s slumping renewable-energy industry should be learnt from, not dismissed.

These are testing times for the renewable-energy boom in China. On 3 August, the head of a Chinese solar-energy company leapt to his death because his firm couldn’t pay off a loan. Last month, LDK Solar in Xinyu City cut its earnings forecast for the second consecutive time, and watched its stock drop to one-quarter of what it was a year ago after it defaulted on payments. Suntech Power in Wuxi, the world’s biggest producer of solar panels, announced plans to slash production and was forced to accept a loan worth US\$32 million from the local government to stay afloat. China’s wind-turbine manufacturers are also being forced to watch sales drop in an overcapacity market.

China — which produces more solar panels and wind turbines than any other country — likes to promote its renewable-energy industry as a glowing symbol of a technology-driven, green future. But, over the past few years, a glut in the international market, a drop in prices and import tariffs introduced by countries including the United States have left companies with over-supply and debt.

Observers have jumped on the downturn with criticisms of how the Chinese government has supported and protected renewable-energy companies. A widely quoted piece in the Shanghai-based newspaper *First Financial Daily* claimed that the slump signalled the “imminent collapse” of the country’s photovoltaics industry. A gloating editorial in *The Wall Street Journal* in August used events in China to warn the US government against support for its own renewables industry.

Strong stuff, but perhaps not unexpected. It is clear that some companies overshot the market, and there are some grounds for claims that China has engaged in unfair trade practices. When the country entered renewable energy, it instituted protective measures such as requiring that 70% of turbines sold in China be produced with domestically made parts, but those measures have been removed. Many renewables firms have failed and can no longer be propped up by supportive local governments.

However, it is foolish to decide that China’s renewables industry has been — or will be — a failure on the basis of its current problems. The scale of some of these problems has been overstated. *The Wall Street Journal*, for instance, blamed decreased revenue at power companies on underperformance at wind farms, when in fact it had more to do with the increasing price of coal.

By whatever means it has been achieved — and whatever else might be said about it — China’s investment in renewables has been a remarkable project. Just seven years after a renewable-energy law threw government support behind the industry, China went from having almost

no stake in the international market to leading the manufacture of solar photovoltaics and wind turbines, in very competitive industries. China has developed know-how and manufacturing bases: it has the engineers, and it is ready to lead the industry into the future.

The country’s targets will carry the industry forward. Some 62 gigawatts of wind power are currently installed — more than in any other country — and the government has set a target of 200 gigawatts by

“It is foolish to decide that China’s renewables industry has been a failure.”

2020. An even bigger difference will come when the domestic demand for solar energy picks up, which it surely will. Until now, nearly all of China’s photovoltaic units have been exported; domestic use has increased but remains at a relatively low 3.1 gigawatts.

The Chinese government is establishing policies that will encourage this. A new renewable-portfolio standard, to be implemented by the end of this year, will force power companies to generate a mandatory proportion of their energy from renewables, with penalties for those that do not. An upgrade of long-distance power lines, to transfer energy from wind farms or megasolar plants in the west of China to the energy-hungry east, was approved last year. Consumers will be forced to share the burden when a surcharge on electricity from renewables kicks in.

Chinese renewable energy has certainly hit a low point. Many of the 80 or so companies that produce wind turbines will probably have to close. But what *The Wall Street Journal* called a waste of time and money can be seen as healthy competition in an immature market. As China’s renewables industry reorganizes and restructures itself, there may be reasons not to emulate it. But the government, for reasons relating to pollution, climate change and energy security, is firmly behind the industry. And it has built itself a solid platform from which to push on. ■

Life sciences

Survivors of the 2010 University of Alabama shooting chose not to push for the death penalty.

Amy Bishop, the biologist who murdered three colleagues in cold blood and grievously wounded two others, will spend the rest of her life behind bars after an Alabama court sentenced her last week. The Harvard-trained assistant professor had been denied tenure at the University of Alabama in Huntsville. In February 2010, months after her appeal against the decision failed, Bishop, a mother of four, pulled out a 9-millimetre pistol during a faculty meeting in a tiny conference room.

Without saying a word, Bishop methodically shot down fellow biologists Maria Ragland Davis, Adriel Johnson and department chairman Gopi Podilla. A bullet to the head of colleague Joseph Leahy left him, after many months of recovery, blind in his right eye and partially sighted in his left. Staff assistant Stephanie Monticciolo, the department’s mother hen, had the teeth on one side of her mouth knocked out. She sustained shattered sinuses and a broken jaw, and was blinded in one eye.

“Many, many things are better than I could have ever hoped,” Monticciolo’s adult daughter Michele posted on a blog 18 months later. “Some things, however, will never be the same.”

The same is true in the biology department at Huntsville, two and a half years after the shooting. Yet signs of a determined recovery abound. Ten new graduate students enrolled in August. The department has made university biochemist Debra Moriarity its chairwoman, and has hired two new faculty members. Leahy, a microbiologist, is back teaching full-time as of this term. And last week,

structural biologists on the faculty hosted an international conference on the crystallization of biological macromolecules, attended by more than 200 scientists.

In the state of Alabama, there are only two possible sentences for capital murder, with which Bishop, now 47, was charged: life in prison with no possibility of parole, or death by lethal injection or electrocution. Prosecutors said almost from the outset that they would seek the death penalty.

At first, Bishop pleaded not guilty “by reason of mental disease or defect”. Then, weeks before the trial was set to open last month, a heartening tale of human generosity began to unfold. It emerged that the spouse of one of the murdered biologists had written to judge Alan Mann, who would have the final say over the sentence if Bishop was found guilty. The letter-writer noted that his or her family had suffered greatly, but added that they could see no benefit in the loss of another life. The writer asked Mann to spare Bishop the death penalty.

The letter prompted Bishop to offer, through her lawyers, to change her plea to guilty if the prosecutors would drop their pursuit of the death penalty. The prosecutors sounded out the other survivors: the nine who had been in the conference room at the time of the shooting, and the families of the dead. None wanted the death penalty. A deal was reached and Bishop changed her plea to guilty.

Robert Broussard, the lead prosecutor on the case, told *Nature* this week that the common sentiment among the survivors “absolutely” swayed him not to seek the ultimate punishment. And so, on 24 September, after a brief trial, Bishop was sentenced to life behind bars.

In 25 years of prosecuting murders, Broussard said, he has never seen such equanimity in so many people affected by a violent crime. Those who will spend a lifetime bearing the wounds that Amy Bishop inflicted, inside and out, reached deep and found mercy. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq

C. BERNARD



Gene therapies need new development models

As with other medicines, the approval of gene therapies should hinge on a risk–benefit analysis for the patient, argues Fulvio Mavilio.

Is gene therapy finally becoming a reality? The European Commission is poised to authorize, for the first time in the Western world, the commercialization of a gene-therapy product. Called Glybera (alipogene tiparvovec), it is designed to treat a rare genetic defect involved in fat metabolism.

Success has been a long time coming. Gene therapy was first administered more than 20 years ago, to a child who had a rare disorder of the immune system called adenosine deaminase (ADA) deficiency. Since then, it has struggled to find its place in medicine amid a roller coaster of successes and setbacks, hype and scepticism that has little precedent in modern times. Although the approval of Glybera is a positive move, it is unlikely to herald a new age of gene therapies — not without significant changes to the system. It is no coincidence that no gene therapy has yet been approved in the United States and that no other gene-therapy product is being considered by regulators in Europe.

Here is why. The design, development and manufacture of products such as Glybera — a virus engineered to carry a correct copy of the defective gene — is complex and done mostly in academic centres. Yet legislation introduced in the past decade in Europe and the United States demands that these products be produced under the same rules that cover conventional drugs, in establishments operated with industry-like standards and certified by government agencies.

This is a formidable challenge for academic centres, which tend to lack the necessary human and financial resources. So why is the development of gene therapy focused there, and not in industry, which seems better suited?

The first reason is the financial uncertainty generated by the complex, confused and poorly harmonized regulatory environment — as the history of Glybera shows. At first, the application for its authorization received a negative opinion from two committees at the European Medicines Agency (EMA): the Committee for Advanced Therapies (CAT) and the Committee for Human Medicinal Products for Human Use (CHMP). Only when another body, the Standing Committee of the European Commission, asked the EMA to reconsider the application in a restricted indication did the CHMP eventually recommend approval under “exceptional circumstances”, requiring post-marketing studies and the set-up of a restricted-access programme. The Dutch firm Amsterdam Molecular Therapeutics, the inventor of Glybera, did not survive the process, and became known as uniQure after refinancing.

Lack of resources is a second reason. For many years, the drug industry stayed away from gene therapy, perceiving it as a dangerous technology of dubious efficacy that was too complex to develop and targeted too small a market.

There are some positive signs, because this last

perception, at least, is changing: the industry now recognizes that rare diseases and orphan-drug legislation provide attractive opportunities. Some recombinant proteins and monoclonal antibodies originally developed as orphan drugs have been repurposed for larger indications.

An example of how academia and industry could cooperate comes from the recent alliance between the drug giant GlaxoSmithKline (GSK) in London, and the charity-funded San Raffaele Telethon Institute for Gene Therapy (TIGET) in Milan, Italy. GSK gained an exclusive licence to develop and commercialize the ADA treatment, and will co-develop with TIGET gene therapies for six more genetic diseases. The contribution of public or charity-funded organizations in early development phases lowers the cost and risk of investing in diseases with a tiny market, and gives the industry access to technologies that can be expanded to more profitable applications, thereby repaying the investment

and allowing resources to be fed back into rare diseases. Unfortunately, promising therapies for hundreds of orphan diseases are unlikely to attract similar industrial interest.

So, how do we ensure that scientists will continue to develop such treatments? Should they all turn to the ‘hospital exemption’, which permits experimental therapies to be manufactured and used under the responsibility of a physician without regulatory supervision?

That should not become standard practice. Governments, funding agencies, scientists and patients’ associations must together come up with new models. Public funds could be used to pay for centralized manufacturing facilities or to subsidize enterprises with the necessary experience

to get involved, as is done for vaccines. And regulators should look again at product definition and the pathway to market.

The complex combination that forms the basis of the ADA therapy makes it somewhere between a biotherapeutic and a transplantable organ, so it hardly meets the definition of a ‘medicinal product’ and should therefore be regulated differently. Moreover, the amount of preclinical data and post-treatment monitoring currently required to authorize a treatment is hardly justified when preclinical models are uninformative and patients have no therapeutic alternatives.

The major factor in deciding whether to authorize an experimental treatment should be a risk–benefit analysis for the patients. Applying a different standard to gene therapy is unfair, slows down its development, discourages investment and ultimately denies people the right to have timely access to possible cures. ■

Fulvio Mavilio is scientific director of Genethon in Evry, France, and a professor of molecular biology at the University of Modena and Reggio Emilia in Italy.
e-mail: fmavilio@genethon.fr

THE INDUSTRY NOW
RECOGNIZES
THAT RARE DISEASES
AND ORPHAN-DRUG
LEGISLATION PROVIDE
ATTRACTIVE
OPPORTUNITIES.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/wuwlfc

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ASTRONOMY

Missing Galactic baryons spotted

More than half of the expected number of baryons — subatomic particles that make up our everyday world, such as protons and neutrons — in the Milky Way are unaccounted for, but Anjali Gupta of Ohio State University in Columbus and her colleagues seem to have spotted the missing matter. They analysed X-ray satellite data and found evidence of a hot gas halo around the Galaxy extending for 100 kiloparsecs. The authors estimate the mass of the halo to be between 10 billion and 60 billion times the mass of the Sun. The cloud could account for the missing particles, they conclude.

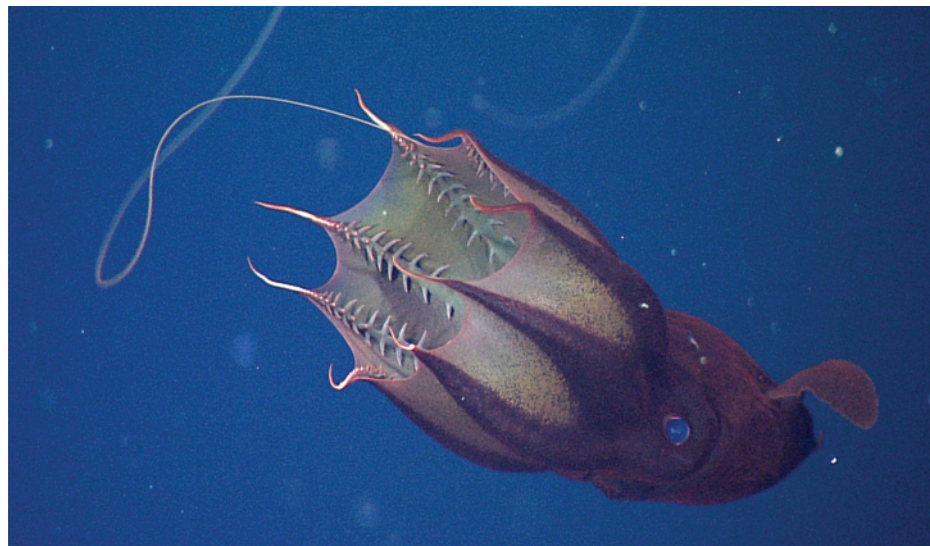
Astrophys. J. Lett. 756, L8 (2012)

ZOOLOGY

Mother's stress slows learning

Female sticklebacks that are confronted by predators while producing eggs generate offspring with impaired learning abilities.

Katie McGhee of the University of Illinois at Urbana-Champaign and her colleagues used a fake predator to repeatedly chase one set of female threespined sticklebacks (*Gasterosteus aculeatus*; pictured), while leaving another set to produce their eggs in peace. Adult offspring from both groups



ANIMAL BEHAVIOUR

Feeding habits of the vampire squid

Drifting in the deep ocean, the vampire squid (*Vampyroteuthis infernalis*; pictured) has features of both octopuses and squid. Researchers have now worked out what and how this mysterious creature eats.

V. infernalis — which is related to octopuses and squid — has eight arms and, instead of the feeding tentacles used by squid to capture prey, has two long, retractile filaments. Hendrik Hoving and Bruce Robison at the Monterey Bay Aquarium Research Institute in Moss Landing, California, studied the feeding behaviour of

V. infernalis using deep-sea video recordings, lab feeding studies and morphological examinations. They conclude that the filaments help the animals to capture food, which includes zooplankton, crustacean remains and even faeces.

The filaments are homologous to the arms of octopuses and other cephalopods, although the creatures' feeding habits are very different, the authors suggest.

Proc. R. Soc. B <http://dx.doi.org/10.1098/rspb.2012.1357> (2012)

NEUROSCIENCE

Responses vary in autistic brains

Autism may emerge from a general unreliability of neuronal responses in the brain's cortex, rather than from a deficiency in one particular brain area or circuit.

Ilan Dinstein at Carnegie Mellon University in Pittsburgh, Pennsylvania, and his colleagues studied 14 people with autism and 14 people without the disorder. The researchers

used functional magnetic resonance imaging to monitor cortical responsiveness as they stimulated the participants' sight, hearing and touch in dozens of trials.

Responses in the visual, auditory and somatosensory areas of the cortex all varied much more between trials for the volunteers with autism than for the controls. The authors propose that this may reflect inappropriate development of neuronal connections, or synapses, in the autistic brain.

Neuron 75, 981–991 (2012)

initially showed similar performances in a task in which the animals learned to associate the colour blue with a food reward. But after five days of the task, the offspring of mothers exposed to predators took twice as long to find the food as did the control group.

At least in these fish, maternal stress can have long-lasting effects on the learning ability of offspring, the authors say.

Biol. Lett. <http://dx.doi.org/10.1098/rsbl.2012.0685> (2012)



B. WATTS/ALAMY

MBARI

ATHEROSCLEROSIS

Fatty plaque link to inflammation

A major constituent of the fatty arterial plaques seen in heart disease may dampen the inflammation that drives the disease. A cholesterol-precursor molecule may be a mediator of this suppression.

Atherosclerosis progresses as cholesterol-filled immune cells called macrophage foam cells accumulate in arterial walls. Christopher Glass at the University of California, San Diego, and his colleagues found that in mice fed a high-fat, high-cholesterol diet, these cells were linked to suppression of inflammation. Within the foam cells, desmosterol, an intermediate in the cholesterol biosynthesis pathway, accumulated to significantly higher levels than did other intermediates. Moreover, desmosterol was also abundant in human atherosclerotic lesions. In the mice, desmosterol seemed to suppress genes that promote inflammation during foam-cell formation.

Synthetic formulations of desmosterol might intervene in cardiovascular disease, the authors suggest.

Cell 151, 138–152 (2012)

ASTRONOMY

Homing in on a black hole's jets

Astronomers have for the first time located the launch site of a giant, high-speed jet of

charged particles believed to originate from a supermassive black hole.

Black holes suck up large amounts of gas and dust, which swirl around the black hole and are thought to feed these jets (simulated image pictured). Sheperd Doeleman of the Massachusetts Institute of Technology's Haystack Observatory in Westford and his colleagues linked four radio dishes in California, Arizona and Hawaii to make a single large telescope. Using this, they examined the jets' source region, which lies just outside the 6.2-billion-solar-mass black hole at the centre of the galaxy M87.

The launch site was small, suggesting two key properties about the gravitational giant: the black hole spins, and it feeds on a surrounding disk of material that orbits it in the same direction as its spin.

Science <http://dx.doi.org/10.1126/science.1224768> (2012)

For a longer story on this research, see go.nature.com/tqrwgb

CELL BIOLOGY

A way to catch dividing cells

A genetic tool can help biologists to pinpoint rare, replicating cells in tissue samples from adult mice.

Finding and studying dividing cells is important in learning about growth, healing and other key processes, but researchers have been hindered by the difficulty of isolating these cells alive. To overcome this problem, Amir Eden at the Hebrew University of Jerusalem and his colleagues created a mouse strain in which the gene for a fluorescent protein was fused to a gene that is active when cells divide. After dissecting livers from these mice, the researchers could sort dividing from non-dividing cells and compare their gene expression. Genes

COMMUNITY CHOICE

The most viewed papers in science

ELECTRONICS

Bendable battery yields flexible LED

HIGHLY READ
on pubs.acs.org
in August

Thanks to an energy-dense flexible lithium-ion battery, researchers have built a thin, bendable light source using an organic light-emitting diode (LED).

The device could one day be incorporated into rollable or implantable electronics.

Lithium-ion batteries are among the best candidates for flexible power sources, but their electrodes could previously be made with only a few, low-performing materials. Now, Keon Jae Lee at the Korea Advanced Institute of Science and Technology in Daejeon and his colleagues have created a bendable battery by using existing methods to apply lithium-based electrodes onto a brittle mica surface at high temperature and then peel off the mica substrate. They then used a technique that they had devised to transfer the battery onto a flexible polymer. The method allows the incorporation into a flexible battery of almost all the high-performance materials that are used in rigid batteries.

Nano Lett. 12, 4810–4816 (2012)

for liver specialization, or differentiation, were less active in replicating cells.

This fluorescent marker could help scientists to isolate and study dividing cells in multiple tissues and biological assays.

Dev. Cell <http://dx.doi.org/10.1016/j.devcel.2012.08.009> (2012)

NEUROSCIENCE

Zebrafish find light without eyes

Eyeless zebrafish larvae may still find their way out of darkness by the activation of light-sensitive neurons deep inside the brain.

Harold Burgess at the National Institute of Child Health and Human Development in Bethesda, Maryland, and his team found that the transparent zebrafish larvae (*Danio rerio*; pictured, with certain brain cells in green) swim gradually towards the illuminated areas of their tank, even after their eyes have been removed. This behaviour suggests the

presence of light-responsive neurons outside the conventional visual organs. Engineered eyeless fish that express less Opn4a — a light-sensitive molecule in the brain — responded poorly to light, whereas fish that produced more Opn4a performed better.

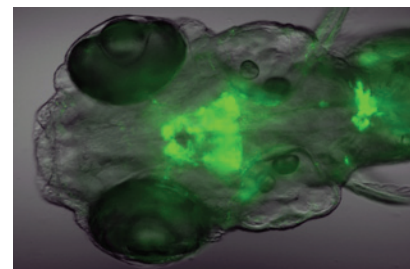
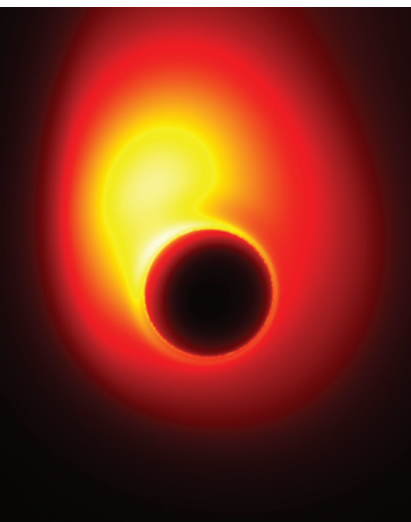
The researchers conclude that neurons expressing Opn4a in the preoptic area of the brain may support simple, and perhaps primordial, light-seeking behaviours.

Curr. Biol. <http://dx.doi.org/10.1016/j.cub.2012.08.016> (2012)

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch



SEVEN DAYS

The news in brief

POLICY

French science safe

France's research and higher-education ministry received a 2.2% boost in the government's 2013 austerity budget on 28 September, escaping the cuts imposed on many other ministries. The increase will pay for 1,000 new university posts. Funding for research grants will also rise by 1.2% to €7.86 billion (US\$10.1 billion) — a cut in real terms if inflation averages above 1.75% as expected. See go.nature.com/7ialrr for more.

Campuses struggle

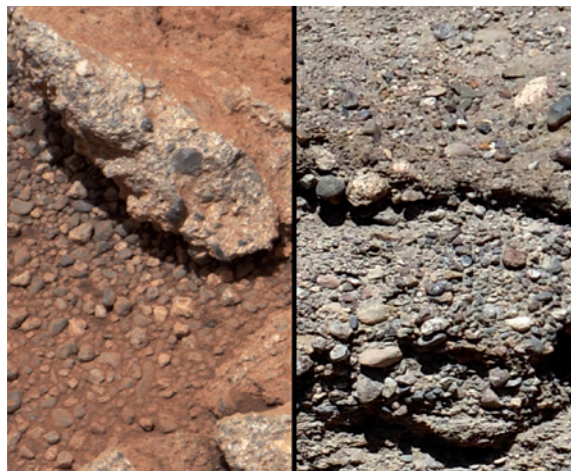
State funding for US public research universities has been insufficient to keep pace with rising student enrolment over the past decade, according to a report by the National Science Board. Between 2002 and 2010, state funding per student dropped in 43 of the 50 states, with cuts as high as 48%. The board warns that the trend threatens to severely hinder research and development at the nation's 101 major public institutions, which train the majority of US scientists and engineers.

Spanish pain

Spain's research and development budget will be cut for a fourth year if the draft budget for 2013 is approved. Presented to parliament on 29 September, the draft budget reduces science spending to €5.9 billion (US\$7.6 billion) — €461 million or 7.2% less than in 2012. See go.nature.com/msiepf for more.

US cuts loom

Federal research and development funds in the United States could be slashed by US\$57.5 billion over the next five years under an across-the-board budget cut that is due to come into effect



Ancient stream on Mars

NASA's Mars rover Curiosity has discovered evidence that water flowed at the bottom of Gale Crater billions of years ago. Although scientists have found many hints of water on Mars before this discovery, rounded gravel pieces photographed by Curiosity (left) are similar to rocky outcrops on Earth (right). These suggest that the stream coursed at speeds of around one metre per second and was at least ankle-deep. See go.nature.com/fyogfs for more.

on 2 January 2013, according to an analysis released on 27 September. The analysis, by the American Association for the Advancement of Science based in Washington DC, predicts that the National Institutes of Health stands to lose US\$11.3 billion, or 7.6%, from its R&D budget as a result of the 'sequester' unless Congress can agree on an alternative budget plan to lower the federal deficit. See go.nature.com/jjrly8 for more.

Nuclear concerns

Plans to build a controversial type of uranium enrichment plant were given the green light by the US Nuclear Regulatory Commission on 25 September. Critics fear that the technology to be used at the facility in Wilmington, North Carolina, proposed

by General Electric–Hitachi Global Laser Enrichment, could lead to the proliferation of nuclear weapons. See page 5 for more.

Fusion failure

The National Ignition Facility, a US\$3.5-billion laser fusion facility at Lawrence Livermore National Laboratory in Livermore, California, did not meet a 30 September congressional deadline for 'ignition' (see *Nature* **483**, 133–134; 2012). That is the point at which the energy produced by the fusion process surpasses that put into the laser shot to trigger fusion.

Mars mission plan

NASA's Mars programme should retain a focus on searching for evidence of past life, rather than on looking for

life today, according to a study released on 25 September. The report, by NASA's Mars Program Planning Group, comes in the wake of the agency's withdrawal from international missions in 2016 and 2018. NASA's next Mars mission is likely to be either an orbiter in 2018 or a rover in 2020, the study suggests. See go.nature.com/wldhwn for more.

Saved species list

At its conference in Jeju, South Korea, earlier this month, the International Union for the Conservation of Nature (IUCN) agreed to produce a 'green list' of fully conserved species. The news was announced last week by the Wildlife Conservation Society, based in New York, who co-sponsored the motion. The list will serve as a counterpoint to the IUCN's 'red list' of endangered species. See go.nature.com/tqhvdb for more.

BUSINESS

Trials monitor

In an announcement on 26 September, the US Department of Health and Human Services charged the Food and Drug Administration with monitoring whether data for clinical trials of drugs and medical devices are incomplete, false or misleading, and notifying the companies responsible. It is unclear, however, what the penalties will be for companies that fail to comply. See go.nature.com/tv24eb for more.

RESEARCH

Coronavirus cases

Scientists in the Netherlands have deposited the full sequence of a new coronavirus that is thought to have caused a respiratory illness in a person

MSSS AND PSI/JPL-CALTECH/NASA

from Saudi Arabia who died from the infection in June (GenBank accession number JX869059; see go.nature.com/g343qd). The sequence matches the partial sequence of a virus from another patient with similar symptoms who was transferred from intensive care in Qatar to London in early September. See page 20 for more.

Ape habitat shrinks

The first continent-wide survey of African great-ape habitat has reported a massive decline between 1995 and 2010. A paper published on 23 September (J. Junker *et al. Divers. Distrib.* <http://doi.org/jfv>; 2012) shows that Cross River gorillas (*Gorilla gorilla diehli*) have experienced a 59% loss in their habitat during this period; bonobos (*Pan paniscus*) 29%; and central chimpanzees (*Pan troglodytes troglodytes*) 17%. See go.nature.com/brilxf for more.

Bas-Congo virus

A virus that causes fever and bleeding and that killed two teenagers and infected a nurse in the Democratic Republic of Congo in 2009 has been identified as a new form of rhabdovirus (pictured). According to a paper published on 27 September, Bas-Congo virus is from the same family as rabies, and



antibody tests suggest that it could be transmissible between humans (G. Grard *et al. PLoS Pathog.* **8**, e1002924; 2012).

Comet discovery

The discovery of comet C/2012 S1 (ISON) was announced by the International Astronomical Union's Central Bureau for Astronomical Telegrams in Cambridge, Massachusetts, on 24 September. The comet is named after the International Scientific Optical Network, which includes the Russia-based telescope that spotted it. Some astronomers suggest that the comet might be bright enough to be seen during the day at around the time it brushes past the Sun in November 2013; however, such early predictions often prove unreliable.

Element 113

Researchers in Japan have made their third atom of element 113 — a feat that

could give them the right to name it. Russian and US researchers may already have created atoms of the element in earlier experiments, but their attempts failed to satisfy the body of experts responsible for deciding on the matter. The experts, drawn from the International Union of Pure and Applied Chemistry and the International Union of Pure and Applied Physics, have yet to report on the Japanese claim. See go.nature.com/cxj91x for more.

PEOPLE

Ecologist dies

Pioneering environmentalist, ecologist and one time US presidential candidate Barry Commoner died on 30 September, aged 95. The scientist rallied against poverty, pollution and nuclear testing and had a key role in the first Earth Day in 1970. He ran unsuccessfully for president in 1980, garnering 234,000 votes.

Chemist accused

A Russian chemist accused of aiding attempted drug trafficking is reportedly facing fresh charges. Olga Zelenina, a narcotics expert at the Penza Agricultural Institute, was released on 25 September from pre-trial detention. But according to Russian news reports, she has since been

COMING UP

7 OCTOBER

Private spaceflight firm SpaceX, based in Hawthorne, California, is hoping to launch its first NASA-contracted cargo-resupply mission to the International Space Station. www.spacex.com

8–10 OCTOBER

The winners of the 2012 Nobel prizes in physiology or medicine, physics and chemistry are announced in Stockholm. www.nobelprize.org

accused by the Russian Federal Drug Control Service of having produced her report on the amount of opiates in a seized shipment of Spanish opium poppy seeds without the necessary permission from her institute — even though the report bears the signature of her institute director, Alexander Smirnov. Smirnov has not responded to *Nature's* request for clarification.

Polar expert cleared

Charles Monnett, a researcher at the US interior department's Bureau of Ocean Energy Management, headquartered in Washington DC, has been cleared of scientific misconduct after a government investigation. Monnett had been accused of publishing false data in a paper that suggested four drowned polar bears had died while swimming in search of sea ice (C. Monnett and J. S. Gleason *Polar Biol.* **29**, 681–687; 2006). He was, however, reprimanded for leaking government documents that later helped environmental groups to sue the government. See go.nature.com/dfio5o for more.

NATURE.COM

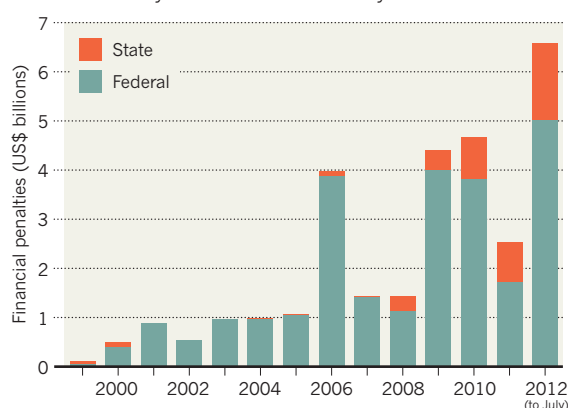
For daily news updates see: www.nature.com/news

TREND WATCH

Pharmaceutical companies paid a record amount in malpractice fines in the United States in the first half of 2012. A report by Public Citizen, a consumer group based in Washington DC, found that by 18 July, US\$5 billion in federal settlements and \$1.6 billion in state settlements had been made for activities such as illegal marketing and overcharging of government medical programmes. Settlements with individual states have steadily increased since 2009.

US PHARMA FINES

Penalties paid out by pharmaceutical companies in the United States have already reached record levels this year.



NEWS IN FOCUS



ASTRONOMY Spy agency bequeaths space telescopes to science **p.16**



INFECTIOUS DISEASE Lab will tackle TB in South African stronghold **p.19**

MISCONDUCT Fraud drives most retractions — but journals are keeping mum **p.21**

ASTRONOMY The rough-and-tumble life cycle of the Milky Way **p.24**

R. BOURGOING/GLOBAL FUND



Under the Affordable Medicines Facility – Malaria programme, shopkeepers sell malaria treatments alongside groceries and over-the-counter medicines.

PUBLIC HEALTH

Malaria plan under scrutiny

Lack of data and donor uncertainty leave public-health experiment on the rocks.

BY AMY MAXMEN

Is showering a country with low-cost malaria drugs the best way to stem the disease? As a US\$463-million pilot programme to test the strategy in seven African countries winds down, public-health experts are questioning whether the approach makes sense given shrinking global health budgets and a steady decline in malaria prevalence.

Although no official decision has been announced about whether to continue the programme, known as the Affordable Medicines Facility – Malaria (AMFm), many of those familiar with it have told *Nature* that it

must change or be phased out after this year.

“For me, the problem is that it has not been proven that the AMFm made a difference to malaria,” says Alan Court, senior adviser to the United Nations special envoy for malaria. “There has to be a public-health purpose or else there is no purpose.” Court is chair of the working group expected to recommend a path for the programme’s future on 14–15 November during a board meeting of the Global Fund to Fight AIDS, Tuberculosis and Malaria, which is based in Geneva, Switzerland, and oversees the AMFm.

➔ **NATURE.COM**

Read more in
Nature's Outlook
on malaria:
go.nature.com/spwwfg

When the AMFm was conceived in 2004, malaria rates in developing countries were skyrocketing, in part because the malaria-causing parasite *Plasmodium falciparum* had become resistant to the drug chloroquine. Officials worried that resistance would also develop to artemisinin — a newer and more effective drug — if people did not combine it with other therapies, but such combinations were much more expensive than either chloroquine or artemisinin alone.

The AMFm aims to make artemisinin-based combination therapies (ACTs) readily available and affordable in malaria-ridden countries by relying on the free market for ▶

► their distribution. The AMFm subsidizes ACTs for drug importers in the pilot countries (Ghana, Kenya, Madagascar, Nigeria, Niger, Uganda and Tanzania), who then distribute the drugs to public and private vendors at low cost. Vendors can sell the drugs to anyone, even people who have not been diagnosed with malaria. The point is to make the drugs accessible to those who would otherwise not get them — such as children and people in rural areas. “I grew up very far from public-health facilities” in Cameroon, says Emmanuel Nfor, acting director of the AMFm, “so to me, the AMFm is not only a good idea, it is an excellent idea”.

Yet, although the AMFm subsidized 60% of the world’s supply of ACTs last year, it is unclear how many of the drugs reached the pilot programme’s target populations. Because anyone can buy the pills, people who can afford them may take them whenever they feel ill, whether or not malaria is the culprit. And although children under the age of five account for 86% of malaria deaths worldwide, the AMFm estimates that in 2011, only 36% of the subsidized ACTs bought by the private sector were for children’s formulations.

The Global Fund commissioned an independent evaluation by ICF International, a consulting firm in Fairfax, Virginia, and the London School of Hygiene and Tropical Medicine (LSHTM), UK, to look at the impact of the AMFm, but it has not settled the debate over the programme’s effectiveness. “We see that the AMFm has been a game-changer in the private for-profit sector,” says Sarah Tougher, a health economist at the LSHTM and part of the committee that evaluated the study (the preliminary report is available at go.nature.com/nsc6ck). According to Tougher, large changes in the price and availability of ACTs, as well as the share of the market supplied by private vendors, “were achieved in just a few months”.

But these benchmarks miss the point, says Mohga Kamal-Yanni, a senior health adviser at Oxfam GB in Oxford, UK. “You don’t need a huge independent evaluation to calculate that a huge subsidy will permit shopkeepers to buy

and sell more of a drug,” she says. “Sales don’t mean anything unless you know who the sales are for.”

Testing before treatment was not considered essential when the AMFm was proposed because easy-to-use diagnostic tests were not available at the time. They are now, and they are also more relevant than they were in 2004 because malaria prevalence has declined

But Albert Peter Okui, a programme manager for the National Malaria Control Programme in Kampala, Uganda, points out that adults could simply take multiple children’s doses if their versions are no longer on store shelves.

Another option is to reduce subsidies in countries with smaller burdens of malaria. Court says that this could allow for an expansion of the programme beyond the pilot

countries, which were selected on the basis not only of malaria burden but also because private shops were heavily involved in drug distribution and because the countries had handled prior Global Fund grants effectively. An expanded programme could potentially reach countries such as the Democratic Republic of Congo, which in 2010 has more reported cases of malaria than four of the pilot countries — Ghana, Madagascar, Niger and Nigeria — combined (see ‘Hot zones’).

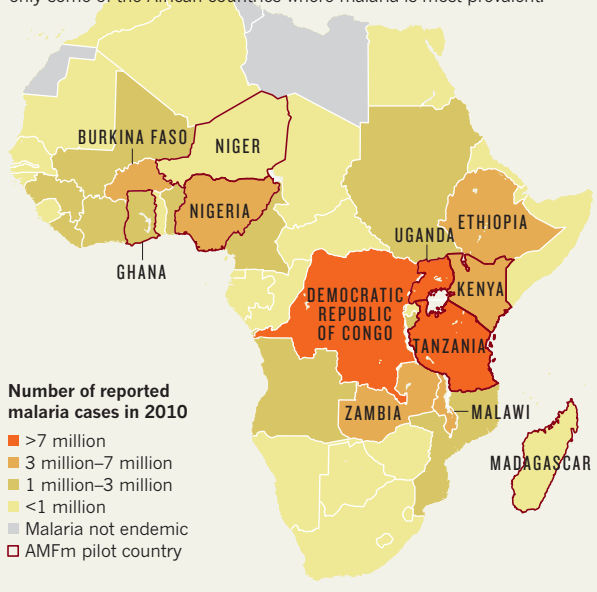
The United States, represented by the President’s Malaria Initiative, has a seat in the AMFm working group. In a document posted online on 28 September, the initiative states its concerns with the programme, including the overuse of ACTs by people who did not need them. The United States did not support the AMFm pilot directly because officials questioned whether a top-down subsidy to importers would get drugs to the most vulnerable groups.

Ultimately, donors will dictate whether the AMFm marches on, says Oliver Sabot, executive vice-president at the Clinton Health Access Initiative in Boston, Massachusetts, and a member of the AMFm working group. “People are committed to working their hearts out to find technical solutions,” he says, “but if there’s not enough funding we will need to roll back the whole thing.”

So much uncertainty just weeks before the decision is nerve-racking for officials in the pilot countries and for drug importers, hospitals and small pharmaceutical companies that depend on the subsidies. “We should be told about this,” says Sigsbert Mkude, a programme officer at the National Malaria Control Programme in Dar es Salaam, Tanzania. “We have little information about what comes next.” ■

HOT ZONES

The Affordable Medicines Facility – Malaria (AMFm) project includes only some of the African countries where malaria is most prevalent.



since then. Today, in some AMFm-supported nations, a child with a fever is as likely to have pneumonia or another fever-inducing ailment as to have malaria.

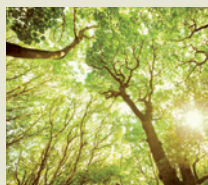
With no way to know how much of the subsidized drugs have been wasted, leaders in global health will struggle to convince donors that the programme is cost effective. No one has yet stepped forward to fund the next phase, and the AMFm working group is debating modifications and exit strategies. “We are grappling with the cost of all of this,” Court says. “For example, we know that the vast majority of people who die of malaria are children; should we just target them?”

In one future scenario, a transformed AMFm would subsidize paediatric doses, which cost less than the adult formulations.



**MORE
ONLINE**

TOP STORY

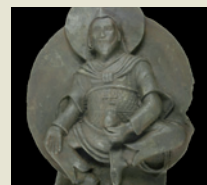


Nitrogen limits plants’ ability to mop up our excess CO₂
go.nature.com/fej8wk

MORE NEWS

- Ice may lurk in shadows far from the Moon’s poles go.nature.com/3idgzp
- Tiny fossils hint at the effects of ocean acidification go.nature.com/ist7wb
- Biodegradable electronics are here today, gone tomorrow go.nature.com/ioscor

FROM THE BLOG



Buddhist ‘Iron Man’ found by Nazis is from space
go.nature.com/79578o



Progeria is one condition that has benefited from a redeployment of 'failed' drugs.

DRUG DEVELOPMENT

New cures sought from old drugs

Researchers to re-examine compounds shelved by industry.

BY MEREDITH WADMAN

Progeria is a rare, lethal disease that ages children so rapidly that they seem to be 80 years old when they are just 10. So patients and families celebrated the news last week of the first therapeutic success against the disease — as did advocates of the notion that abandoned drugs can rise again. A clinical trial has shown improved symptoms in children with the disease who used lonafarnib (L. B. Gordon *Proc. Natl Acad. Sci. USA* <http://doi.org/jfz; 2012>), a drug developed by Merck in the 1990s that failed against its original target, head and neck cancer.

If resurrecting an ineffective drug works for progeria, why not for other conditions, asks Francis Collins, director of the National Institutes of Health (NIH) in Bethesda, Maryland, whose lab in 2003 identified the mutated gene that causes progeria. “And why not seek those answers by ‘crowd-sourcing’?” he asks.

Roughly 30,000 drugs have been shelved by the pharmaceutical industry over the past three decades, and medical funders are now inviting researchers to find a new future for some of them. This month, the UK Medical Research Council is expected to announce the

initial awards in a £10-million (US\$16-million) programme aimed at repurposing 22 stalled compounds developed by London-based AstraZeneca. And this week, the NIH’s National Center for Advancing Translational Sciences (NCATS) will solicit the first full applications to a similar programme it launched in May.

NCATS invited researchers to look for new uses for 58 abandoned compounds contributed by eight big drug companies. All were taken through preclinical testing and into early human trials at a cost of millions of dollars, and all were found to be safe. They were abandoned either for lack of efficacy or for business reasons.

The planned budget for the NCATS programme is just \$20 million in 2013, but the competition seems to be fierce: by mid-August, the agency had received nearly 160 pre-applications for a maximum of eight awards. Kathy Hudson, acting deputy director of NCATS, says that the pre-applications sought to attack conditions as diverse as autism, Alzheimer’s disease and cancer. For many compounds, multiple uses

were conceived; for one compound alone, the agency received seven applications targeting six diseases or conditions.

➔ NATURE.COM
For details on the NCATS drugs, see: go.nature.com/u4li4r

This week, the most promising will be asked to submit full applications.

The winners, to be announced in June 2013, will have to provide preclinical evidence of the biological rationale for the compound’s new use. The company then gets to decide whether it wants to take that finding any further.

John LaMattina, former president of global research and development at Pfizer in New York, thinks that the programme faces long odds because drug companies have already tested these compounds against many targets. “There has been an awful lot of hype,” says LaMattina, who is now a senior partner at PureTech Ventures, a life-sciences venture-capital company in Boston, Massachusetts. Any new finding by an NIH grant recipient, he adds “is going to have to be pretty compelling for a company to take it forward”.

Steven Potkin, a psychiatrist at the University of California, Irvine, worries that even compelling findings may not be enough to persuade companies to reinvest in a shelved compound. Potkin wants to repurpose a compound originally aimed at depression that he thinks could help with mood-regulation problems across a range of psychiatric diseases from bipolar disorder to schizophrenia. But the NIH requires the participating companies to share the intellectual property created if grant recipients find new uses for a compound; a company can then license it back to develop the drug. “There is no guarantee that this will happen,” Potkin says.

Hudson says the companies have agreed that, if they are not interested, the academic partner can either purchase the agent from the company, or have it manufactured by a third party, so that the push towards the clinic can still proceed. Then, “if the company elects not to pursue the new indication, all is not lost”, Hudson says.

Chemists are grumbling about a different issue: the companies have not released the structures of the 58 compounds (although the programme’s winners will learn the structures

“This project has required a delicate dance to get industry to share compounds and data.”

of their compounds). That has led to a “tremendous waste of effort” for chemists trying to deduce any new activities that the compounds might have, says Jeremy Berg, former director

of the NIH’s National Institute of General Medical Sciences in Bethesda, and now a computational and systems biologist at the University of Pittsburgh in Pennsylvania.

But some say that the NIH had to meet the industry halfway. “This project has required a delicate dance to get industry to share compounds and data,” said Thomas Insel, who was acting director of NCATS until 23 September, in an August e-mail to Berg. “My hope is that if this first step goes well, we can expect more sharing in the future.” ■

ASTRONOMY

The telescopes that came in from the cold

Twin spy telescopes could drive US space astronomy forward, but at what cost?

BY ERIC HAND

When astronomer Alan Dressler was invited to see what might be the future of NASA's astrophysics programme, he had to leave his mobile phone behind, lest he be tempted to grab a quick snapshot. As he and a dozen others passed through the ITT Exelis facility in Rochester, New York, a guide held up a flashing red light, to warn working Exelis engineers to seal their lips in front of people without security clearance.

Their destination was the cavernous clean area of Building 1230, where two 2.4-metre telescopes, each as big as the Hubble Space Telescope and never flown, rested on low pedestals. "It seemed almost too good to be true," says Dressler, an astronomer at the Carnegie Observatories in Pasadena, California. "Things like this just don't drop on your doorstep."

The unexpected gift comes from the US National Reconnaissance Office (NRO), a secretive surveillance agency that built the telescopes to peer down on Earth. In June, NASA revealed that the NRO had bequeathed the scopes to the space agency because they were no longer needed. Now NASA has to figure out what it will do with them — and whether it can afford the cost of kitting them out with instruments and sending them into orbit.

This month, NASA plans to announce a science-definition team that will embark on that assessment. The team will report by April 2013 to NASA administrator Charles Bolden on the pros, cons and costs of adapting one of the telescopes for a mission to investigate dark energy, the phenomenon thought to be accelerating the expansion of the Universe. But astronomers are already encouraged. As the veil of secrecy surrounding the telescopes lifts, astronomers are beginning to size up the devices' capabilities. And so far, they are liking what they see — so much so that they are now talking about tacking on an instrument that would detect extrasolar planets directly. "I think the enthusiasm has only increased as time has gone on," says Dressler.

The most likely first use for an NRO telescope is as an alternative to the proposed Wide-Field

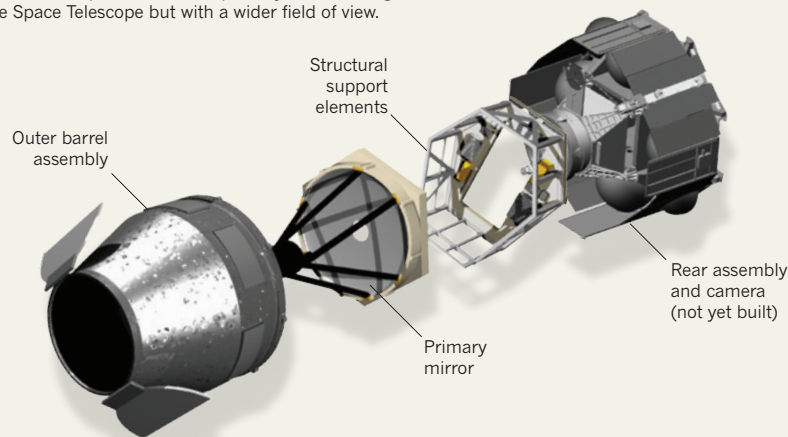
➔ **NATURE.COM**

For more on
space telescopes
see:

go.nature.com/vy7sy8

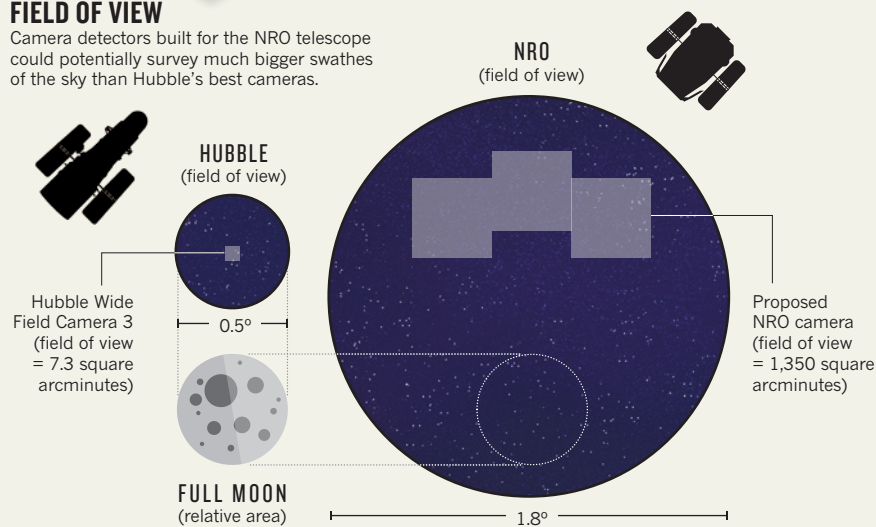
ANATOMY OF A GIFT

The National Reconnaissance Office (NRO), a US spy agency, has given NASA two telescopes, each with a primary mirror as big as that in the Hubble Space Telescope but with a wider field of view.



FIELD OF VIEW

Camera detectors built for the NRO telescope could potentially survey much bigger swathes of the sky than Hubble's best cameras.



Infrared Space Telescope (WFIRST), the top-ranked mission in the 2010 astronomy decadal survey. Before the NRO telescopes came along, astronomers were aiming for a wide-field 1.3-metre survey telescope to search for the imprint of dark energy, to find exoplanets and to study star-forming regions of the Galaxy. WFIRST was not expected to fly until the mid 2020s. But with a telescope already in hand, the NRO version of a WFIRST mission could conceivably launch at the end of the decade — potentially challenging Europe's Euclid space telescope, a

dark-energy mission scheduled to fly in 2019.

The NRO telescope suits the WFIRST mission because it has a much wider field of view than Hubble (see 'Anatomy of a gift'), which makes it perfect for spotting the thousands of supernovae and millions of galaxies needed to pin down dark energy. And at a workshop held last month at Princeton University in New Jersey, Gary Matthews, the director of astronomy at ITT Exelis, released data showing how smoothly the telescope's mirror had been polished. "The mirror is about

NASA

as good as Hubble's," he says.

The telescope's supporting structure is made of a resin that resists distortions caused by temperature changes, which will help keep the main mirror stable. Active control of a secondary mirror could adjust it to correct for any distortions due to the main mirror, further sharpening the telescope's optics.

"This telescope was clearly designed to produce very stable images," says Matt Mountain, director of the Space Telescope Science Institute in Baltimore, Maryland. And extremely stable images will be crucial for the success of one particular dark-energy technique — weak lensing — which looks for subtle distortions in the shapes of galaxies due to intervening dark matter (see *Nature* **489**, 190–191; 2012).

Astronomers acknowledge that the telescope will not be ideal for studying the most distant galaxies in the Universe, which are visible only in the infrared. That requires a system for cooling the mirror but, unlike the proposed WFIRST instrument, the NRO telescope is designed to work at room temperature. However, its larger mirror will have so much more light-gathering power that it will be able to spot many more faint objects nearer to hand, even if it can't stare as deeply into space.

COSTS AND BENEFITS

Some astronomers, however, are questioning whether the value of the free hardware — each NRO telescope is worth at least US\$250 million — can compensate for the extra costs entailed in going from a 1.3-metre mission to a 2.4-metre mission, which will require a larger rocket and a larger camera. Although the WFIRST mission was expected to cost \$1.5 billion, one NASA estimate puts the NRO option at \$1.75 billion.

But Princeton astronomer David Spergel, who organized the workshop, believes that that figure underestimates the savings to be made by using the NRO scope. Not having to cast and polish a primary mirror avoids a long, labour-intensive process requiring an army of technicians, he says. Spergel thinks that a \$1.6-billion mission is realistic.

He would bump up the cost another \$200 million, however, to add an instrument that could take advantage of the extra light-gathering capability of the NRO telescope. This would be a coronagraph, which can block the light of a star while still revealing the dim glow of orbiting planets. WFIRST was going to search for exoplanets by detecting gravitational-lensing events: distortions in the light of a background star caused by a large planet orbiting another star in the foreground. But that method would not easily allow astronomers to detect planets that orbit close enough to be within a star's habitable zone. Astronomers would also prefer to collect a planet's reflected light directly. Mountain says that a modern coronagraph on NRO-WFIRST might be able to detect a Neptune-sized planet. It could

also survey the dust that shrouds many stars and that could block small planets from view. Knowing how ubiquitous and severe the dust problem is would help exoplanet astronomers work out just how big their next big ask — a mission to directly detect Earth-like planets — has to be.

HUMAN INVOLVEMENT

One way to reduce the cost of the NRO-WFIRST mission for NASA's astrophysics division would be to launch it on one of the new fleet of rockets that NASA will be eager to test at the end of the decade as it moves beyond the now-grounded space shuttles. But that would involve NASA's human space programme, an option that the science-definition team has been asked to consider. It could mean moving the mission from its intended orbit around the Sun — at a dynamically stable spot known as a Lagrangian point some 1.5 million kilometres beyond Earth's orbit — to a geostationary orbit about 36,000 kilometres above Earth (still much further out than Hubble). The geostationary option would be within reach of a wider variety of rockets — and of potential servicing missions by astronauts.

Although a Lagrangian point is a better spot for astrophysical observations because, for example, it lies outside Earth's radiation belts and has a more constant temperature environment, a geostationary orbit would allow much higher data-transfer rates, which would be particularly important in a survey mission accumulating vast amounts of data.

Astrophysicists believe that they are following the most sensible path for repurposing the NRO hardware, but they are aware that

"It seemed too good to be true. Things like this just don't drop on your doorstep."

other space scientists wouldn't mind a crack at it too. One idea presented at the Princeton workshop was to investigate Earth's aurora and ionosphere by observing the edge of Earth from a vantage point within the orbit of the Moon. This proposal might need special approval, however, as one of the NRO's stipulations for donating the telescopes was that they would not be used to look at Earth. Such a large telescope could also be useful in planetary science, to search for asteroids likely to pass close to Earth or to study the faint objects beyond Neptune's orbit.

But there is a second telescope that could support these ideas, and there are assorted loose parts, including a primary mirror, for a third telescope, testimony to the NRO's lavish funding. Touring the ITT Exelis facility, Dressler was struck by the number of huge chambers designed to test space telescopes under vacuum. This was a place ready to stamp out Hubble-sized telescopes by the dozen, he says. "It makes you a little jealous," he says. "It's kind of neat and kind of sad." ■



Across Africa, the need for improved tuberculosis diagnosis and treatment is growing.

BIOMEDICAL RESEARCH

Foundation opens TB lab in Africa

Howard Hughes Medical Institute sets up shop at ground zero for tuberculosis: South Africa's KwaZulu-Natal.

BY LINDA NORDLING IN DURBAN

Patients huddle in dressing gowns outside the entrance to the King George V Hospital in Durban, South Africa. It is safer out here: fresh air limits the spread of disease. Inside, the sickest lie listless, too tired to get up, or perhaps beyond caring.

This threadbare hospital is on the front line in Durban's war against multidrug-resistant and extensively drug-resistant tuberculosis (TB). Its patients — most of whom also have HIV — undergo gruelling courses of treatment that last for months or years. Those who get better may experience side effects such as hearing loss or psychosis. The unlucky ones die.

For William Bishai, these bleak circumstances offer a source of hope, through an unusual research collaboration. The microbiologist heads the KwaZulu-Natal Research Institute for Tuberculosis and HIV (K-RITH), which officially opens on 9 October in a new building on the campus of the Nelson R. Mandela School of Medicine at the University of KwaZulu-Natal (UKZN) in Durban.

The province of KwaZulu-Natal has one of the highest rates of TB cases in the world: about

3,000 drug-resistant cases are diagnosed each year. Furthermore, 80% of people diagnosed with TB there also have HIV — making the province perhaps the best place in the world to study the lethal interplay between these diseases. The dire need, and the opportunity to develop new treatments and diagnostics, led the Howard Hughes Medical Institute (HHMI) in Chevy Chase, Maryland — one of the richest biomedical research foundations in the world — to establish K-RITH as its first laboratory outside the United States. The HHMI will spend about US\$75 million on the institute over 10 years; the UKZN will contribute another \$10 million.

The seven-storey, \$40-million K-RITH building is close to Durban's main chest clinic, where 15,000 people are screened for TB, HIV and other sexually transmitted diseases every month. This is where people will be recruited to take part in research. "The samples are literally just outside our door," says Bishai.

➔ **NATURE.COM**
Read more in
Nature's Africa
special:
go.nature.com/ylnyfw

Antibiotics that successfully treat TB have been around since the 1940s, but interrupted or

incomplete courses of treatment using first-line drugs such as isoniazid and rifampicin have led to an alarming worldwide increase in infections with resistant strains of TB (T. Dalton *et al.* *Lancet* <http://doi.org/h8r>; 2012). K-RITH will focus on developing better treatments and faster and more accurate diagnostic tests, not only to improve survival, but also to reduce the spread of the disease. In addition, the institute will study the relationship between TB and HIV, which Bishai says is poorly understood. "It is clear that TB makes HIV worse and HIV makes TB worse. But we don't understand the mechanism behind this," he says.

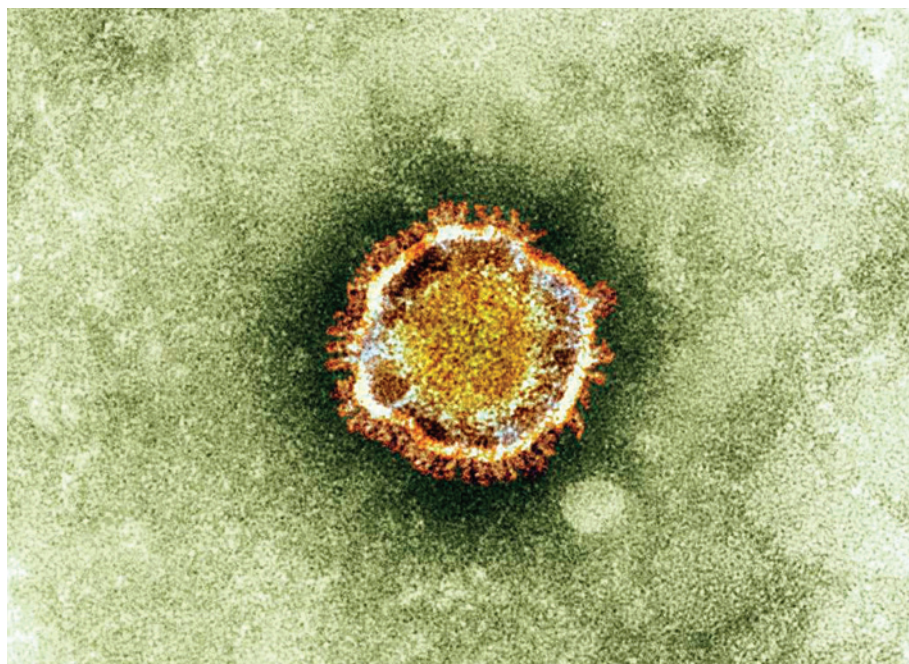
The institute is part of a growing trend in Africa to locate research on diseases close to the people who have them, says Bishai. Samples collected in Africa were once sent to the United States, Europe or Asia for study, but this often leaves local scientists and patients feeling exploited. Co-location of research and patients helps to build scientific capacity in the region, and encourage locals to participate in studies.

Eight investigators already work at K-RITH, and three are African. "We really beat the bushes to find African scientists," says Bishai, who moved from Maryland to Durban last year. Ugandan engineer Frederick Balagaddé will use microfluidic chips — chemistry sets the size of credit cards — to develop improved tests for HIV and TB. Adrie Steyn, a South African microbiologist, will study how *Mycobacterium tuberculosis*, the microbe that causes TB, fends off attacks by the immune system. And Thumbi Ndung'u, a Kenyan virologist, wants to find out why some people are less susceptible to HIV infection than others.

The other investigators are European or from the United States, but all had to agree to spend at least 80% of their time in Durban. One is a legend of TB research: Frenchman Jacques Grosset was diagnosed with TB at the age of 25, and after getting out of the sanatorium he took up the fight against the disease at the Pasteur Institute in Paris. Over the past 50 years he has had a hand in developing most of the drug regimens used to treat TB. Now, at the age of 83, Grosset has moved from Paris to Durban, expecting to finish his long and illustrious career using K-RITH's biocontainment laboratories to study drug-resistant TB in mice.

K-RITH's first clinical study is already under way, and aims to work out the optimum dose of TB drugs to use in children who have both TB and HIV. Another begins recruiting this month, as part of a global trial of a combination of drugs — PA-824, moxifloxacin and pyrazinamide — sponsored by the Global Alliance for TB Drug Development, a public-private partnership in New York (see *Nature* **487**, 413–414; 2012).

Having this hive of research activity on their doorstep is already raising the hopes of local clinicians, says Surie Chinappa, a doctor at the Durban chest clinic. "I think it will be positive for our patients." ■



The research legacy of the SARS virus (pictured) is helping scientists to move quickly against a new threat.

BIOMEDICAL RESEARCH

SARS veterans tackle coronavirus

Genome sequence of new virus speeds up testing.

BY DECLAN BUTLER

Scientists who helped to fight the 2003 epidemic of SARS (severe acute respiratory syndrome) have sprung into action again to investigate the latest threat: a new SARS-related virus that has killed one man and left another seriously ill. Last week, the researchers reported the genome sequence of the new coronavirus and the first diagnostic tests to screen for it — two major advances that will help in efforts to control the pathogen if it turns into a wider menace.

The SARS virus was identified in March 2003 as the cause of an epidemic that had emerged in China several months before, and which had spread rapidly around the world. It caused nearly 8,500 cases and 916 deaths before it was finally contained in July 2003. At the time, scientists knew almost nothing about the virus — coronaviruses had received scant attention until then because they had previously caused little more than colds.

The research and public-health networks established during the SARS epidemic — and the body of coronavirus research that followed — puts scientists today in a much stronger

position to understand the latest virus and to develop countermeasures such as drugs and vaccines, should they be required. “We are all collaborating again,” says Christian Drosten, director of the Institute of Virology at the University of Bonn Medical Centre in Germany, who has been involved in developing diagnostic tests for the pathogen. “This is the old SARS club.”

So far, there is little evidence that the virus poses any major public-health threat. No one who came into contact with the two cases has fallen ill, suggesting that the virus does not spread easily between humans. Nonetheless, health authorities worldwide are not being complacent — respiratory viruses can cause pandemics, and this strain has already caused serious disease. The key question now, and one that the diagnostics will help to answer, is whether the two cases are isolated events or whether the virus could strike again and perhaps adapt to spread more easily in humans.

The first case was a 60-year-old man admitted to the Dr Soliman Fakeeh Hospital in Jeddah, Saudi Arabia, on 13 June with severe pneumonia and acute renal failure, who died on 24 June. Post-mortem tests were negative

for influenza and the other usual suspect viruses, so Ali Mohamed Zaki, a microbiologist at the hospital, ran a coronavirus test at the suggestion of Ron Fouchier, a virologist at Erasmus University Medical Centre in Rotterdam, the Netherlands, who had worked on the SARS virus.

On 20 September, Zaki posted his results on the Program for Monitoring Emerging Diseases (ProMED), an online disease-reporting system, confirming that the coronavirus tests were positive. A week later, Fouchier’s group, which had received an isolate of the virus from Zaki in early July, published the pathogen’s genome sequence in the GenBank database. The genome confirmed that the pathogen was a new coronavirus, the closest relatives of which are found in bats.

Two days after seeing Zaki’s ProMED post, the UK Health Protection Agency reported that it had found a second case: a 49-year-old man from Qatar who fell ill on 3 September with similar symptoms. He was admitted to intensive care in Doha on 7 September and then transferred to a London hospital, where he remains seriously ill. Comparing a fragment of the genome sequence of his virus with that of the first case showed that the two were identical. The viral genome sequence also enabled an international group of researchers, including Fouchier and Drosten, to quickly devise diagnostic tests that look for short, characteristic stretches of the virus’s RNA. The collaboration came easily, says Drosten: “The good thing here is that these are all friends from the SARS time.”

Existing SARS research provides a useful template for further investigation of this latest coronavirus, adds Drosten. Scientists will use animal models such as mice, ferrets and macaques to study the pathogen’s virulence and how it spreads, for example. They will also test whether antivirals and vaccines developed since SARS to treat other coronaviruses are effective.

A key experiment, he says, will be to find where the new virus latches on to the human lung. Some scientists suspect that it might bind to the angiotensin-converting enzyme 2 (ACE2) receptor, as did the SARS virus. That could be both good and bad news. The receptor is found deep in the lungs, where infections can cause severe disease, but viruses nestling there are less apt to be coughed or sneezed into the air than are those found higher in the lungs.

“Receptor-binding properties could also be crucial to the success of potential control measures, should they be needed,” says Drosten. SARS was contained by isolating suspected cases, partly because it did not spread quickly, but also because those it infected became very ill before the virus moved into the upper respiratory system. That made cases easy to identify before the patients started spreading the virus. Flu pandemics, by contrast, are impossible to stop, largely because those infected with the virus spread it to others for days before they show any symptoms of infection. ■

COURTESY OF HPA

PUBLISHING

Misconduct is the main cause of life-sciences retractions

Opaque announcements in journals can hide fraud, study finds.

BY ZOË CORBYN

Conventional wisdom says that most retractions of papers in scientific journals are triggered by unintentional errors. Not so, according to one of the largest-ever studies of retractions. A survey¹ published in *Proceedings of the National Academy of Sciences* has found that two-thirds of retracted life-sciences papers were stricken from the scientific record because of misconduct such as fraud or suspected fraud — and that journals sometimes soft-pedal the reason.

The survey examined all 2,047 articles in the PubMed database that had been marked as retracted by 3 May this year. But rather than taking journals' retraction notices at face value, as previous analyses have done, the study used secondary sources to pin down the reasons for retraction if the notices were incomplete or vague. These sources included investigations by the US Office of Research Integrity, and evidence reported by the blog Retraction Watch.

The analysis revealed that fraud or suspected fraud was responsible for 43% of the retractions. Other types of misconduct — duplicate publication and plagiarism — accounted for 14% and 10% of retractions, respectively. Only 21% of the papers were retracted because of error (see 'Bad copy').

Earlier studies had found that the percentage of retractions attributable to error was 1.5–3 times higher^{2–4}. "The secondary sources give a very different picture," says Arturo Casadevall, a microbiologist at Yeshiva University in New York, and a co-author of the latest study. "Retraction notices are often not accurate."

Elizabeth Wager, a UK-based medical writer and co-author of a previous study³ that relied on journal retraction notices, isn't surprised by the finding of hidden misconduct. "We found many notices that seemed deliberately obscure or vague," she says, speculating that authors and journals may use opaque retraction notices to save face or avoid libel charges.

The latest study shows a ten-fold increase (to about 0.01%) in the proportion of papers retracted owing to fraud since 1975. Previous analyses have seen a growing trend in retractions in general⁵, but the latest report sheds new light on the extent to which fraud is responsible. It also found a correlation between journal impact factor and the number of fraud-induced retractions, says Ferric Fang, a

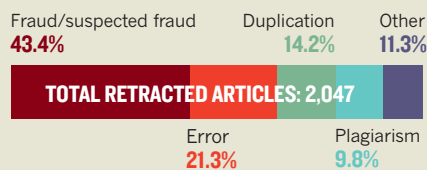
TOP TEN RETRACTORS

Journals with the most retractions attributable to fraud or suspected fraud, as recorded in PubMed.

Journal	Number of articles	2011 impact factor
<i>The Journal of Biological Chemistry</i>	37	5.12
<i>Anesthesia & Analgesia</i>	33	3.07
<i>Science</i>	32	32.45
<i>The Journal of Immunology</i>	30	5.86
<i>Proceedings of the National Academy of Sciences</i>	27	10.47
<i>Blood</i>	21	9.79
<i>Nature</i>	19	36.24
<i>The Journal of Clinical Investigation</i>	17	15.43
<i>Cancer Research</i>	16	8.16
<i>Cell</i>	13	34.77

BAD COPY

Most retracted papers listed in PubMed were withdrawn owing to fraud or suspected fraud.



microbiologist at the University of Washington in Seattle, who led the study.

Influential journals, including *Science*, *Nature*, *Proceedings of the National Academy of Sciences* and *Cell*, all appear in the top-ten list of publications with retractions because of fraud or suspected fraud (see 'Top ten retractors'). For some journals, including the two topping the table — *The Journal of Biological Chemistry* and *Anesthesia & Analgesia* — the tally was boosted by multiple retractions from the same few individuals, such as anaesthesiologist Joachim Boldt, formerly of the Ludwigshafen Clinical Center in Germany. Indeed, Fang and his colleagues found that 38 research groups with five or more retractions accounted for 44% of

articles linked to fraud or suspected fraud.

Whether the overall rise in fraud-induced retractions is the result of an increase in misconduct, or simply down to more scrutiny, is an open question, says Fang. It is also unclear whether the high-impact journals have more retractions for fraud because they are checked more closely, or because they are more likely to attract fraudsters. But Fang thinks that the large rewards for publishing in leading journals — which can range from winning grants to receiving tenure — are powerful incentives that could be driving some of the trend. "We need to look at how we have structured the system, so scientists are not given incentives to [commit fraud] quite as strongly," he says.

The survey found some significant geographical differences. Retracted papers with lead authors based in historical scientific superpowers, such as the United States and Germany, were more likely to be linked to fraud. In emerging scientific powers such as India and China, however, plagiarism and duplication caused more of the retractions. "These trends may reflect differences in incentives, cultural norms and proficiency in English among these countries," says Fang.

Ivan Oransky, a New York-based journalist and co-founder of Retraction Watch, suggests setting up a 'transparency index' for journals, to rank them on criteria such as the clarity of their retraction notices. The idea, which he says he would be keen to work on, could provide a much-needed incentive for journals to improve their performance in this area. Data from the current study could also serve as a basis for a retractions database to help scientists avoid wasting time trying to replicate or build on retracted work, he adds.

"I'm not necessarily opposed to the idea, but I have concerns about how such a database could be properly maintained and updated," says Fang. "Our study is merely a snapshot. Creating an accurate, centralized database that could be used as an ongoing resource would be a considerable undertaking." ■

1. Fang, F. C., Steen, R. G. & Casadevall, A. *Proc. Natl. Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1212247109> (2012).
2. Steen, R. G. *J. Med. Ethics* **37**, 249–253 (2011).
3. Wager, E. & Williams, P. *J. Med. Ethics* **37**, 567–570 (2011).
4. Nath, S. B., Marcus, S. C. & Druss, B. G. *Med. J. Aust.* **185**, 152–154 (2006).
5. Van Noorden, R. *Nature* **478**, 26–28 (2011).

NATURE.COM
Read more in
Nature's retractions
feature:
go.nature.com/cgt4re

COMMENT

HISTORY What do some of the great laboratories of history have in common? **p.31**

DISEASE On the trail of the next human pandemics lurking in animals **p.33**

EXPEDITIONS Among the people behind the Mars rovers **p.34**



PUBLISHING Reviews turn inchoate literature into knowledge **p.37**

A. JOHANSSON / SHUTTERSTOCK.COM



Don't let copyright block data mining

Matthew L. Jockers, Matthew Sag and Jason Schultz explain why humanities scholars have pitched in to the *Authors Guild v. Google* lawsuit.

Advances in computer technology combined with the availability of digital archives are allowing humanities scholars to do what biologists, physicists and economists have been doing for decades — analyse massive amounts of data. A far richer understanding of literature promises to emerge. For instance, large-scale quantitative projects are forcing scholars to reconsider how literary canons are formed and are showing the extent to which authors' works are shaped by factors outside their own creative control, such as the period in which they lived, their gender and their nationality.

Yet in the United States, legal action pursued by the Authors Guild, an advocacy group for writers, could bar scholars from

studying as much as two-thirds of the literary record. A small group of humanities scholars (ourselves included) is fighting back.

CASE HISTORY

In 2004, Google began scanning and digitizing books held in prominent US academic libraries such as those at Stanford University in California and the University of Michigan in Ann Arbor, to make these collections fully searchable. Currently, more than 20 million books, most of which are out of print, can be searched at Google Books (books.google.com). Unless a book's copyright protection has expired, or the copyright owner has agreed to make the content freely available, the search engine

displays just three-line 'snippets' from each book — enough to tell the searcher that the listed item is indeed what they are looking for. With the right tools, however, data from the full text can, in principle, be mined and used in large-scale analyses.

In 2005, the Authors Guild, based in New York, with some 8,500 members including published authors, literary agents and lawyers, filed a class-action lawsuit claiming that Google's scanning activity was a "massive copyright infringement". Google, the Authors Guild and a group of publishers agreed to a class-action settlement in 2008. This gave Google permission to continue scanning and to sell electronic books individually or as part of a subscription service. In return, ►

► Google agreed to share the advertising revenue from Google Books with authors and publishers, and to make one-off payments to copyright owners amounting to a minimum of US\$125 million.

The settlement was strongly opposed by foreign governments, the US Department of Justice, the US Copyright Office, authors, academics and rival technology companies for various reasons. Many feared that it would create an unfair monopoly, with Google having the sole right to publish millions of 'orphan' works — books whose copyright owners cannot easily be located. In 2009, the settlement was revised to try to address these concerns. But the court rejected the revised settlement in 2011, and the legal controversy continues.

In September last year, in a separate case, the Authors Guild sued several universities for participating in Google's book-scanning project. As part of this case, known as *Authors Guild v. HathiTrust*, it is also pursuing legal action against the HathiTrust Digital Library, a service that enables a large consortium of universities and research libraries to store, secure and search their digital collections using a shared infrastructure.

Among the issues at the heart of this dispute is what researchers in the emerging field of digital humanities will be allowed to analyse: only public-domain books (mostly those published before 1923 in the United States), or all known literary works. The answer may define the future of the field.

TO THE BARRICADES

On 3 August, the Association for Computers and the Humanities and a group of 64 scholars (that includes us), from disciplines ranging from law and computer science to linguistics, history and literature, filed an

amicus curiae brief on behalf of the digital humanities. We are urging the court in *Authors Guild v. Google* to grant a summary judgment in favour of Google, a step that will effectively end the litigation¹. We filed a similar brief in the HathiTrust case on 7 July. The judge in the HathiTrust case is currently considering our submission, and a decision is expected imminently. The court in *Authors Guild v. Google* will consider our argument as soon as the appeals court deals with certain procedural issues.

We feel that if the Authors Guild wins the cases against Google and the HathiTrust, the ruling could set a dangerous precedent — that copyright gives authors and publishers the right to control all, even 'non-expressive' uses of their works that involve copying. Copyright law has long recognized the distinction between protecting an author's original expression and the public's right to access the facts and ideas contained within that expression. According to the US Constitution, the purpose of copyright is "To promote the Progress of Science and useful Arts". Preventing authors from monopolizing facts and ideas allows others to explore their own creativity and 'stand on the shoulders of giants'.

We believe that copyright law is not (and should not be) an obstacle to statistical and computational analysis of the millions of books owned by university libraries. We are not talking about republishing them or even quoting from them. We simply want to extract information from and about them to sift out trends and patterns.

As an example, clustering more than 3,000 nineteenth-century novels according to how much they share certain stylistic properties (specific words and punctuation marks) and thematic features (such as groups of commonly

co-occurring words) has thrown up findings that would be hard to glean from reading a handful of books individually. One is that books authored by men tend to cluster quite distinctly from books authored by women (see 'Knowing your subject'). This illustrates the degree to which gender determines the choices made by writers, but also flags up outliers. For instance, within this clustering, the works of George Eliot (real name Mary Anne Evans) sit firmly among those of male writers. In other words, such 'macroanalytic' methodology gives researchers a way to see individual authors and publications within the context of a much larger system.

Authors' rights deserve protection. And governments and the various stakeholders involved may eventually work out how to achieve the full potential of digital libraries in a way that is fair to writers, readers and providers. But digitizing books for 'non-expressive' uses, such as basic searching and text mining, is a separate issue and should not be barred on the basis of concerns over copyright. An independent review last year of intellectual property and growth commissioned by the British government came to a similar conclusion². Unauthorized music-file sharing can infringe copyright because humans ultimately experience those files as musical works. Scanning words from library books to make a search index, or to compile a list of word frequencies, does not interfere with the rights of the author. These uses simply convert masses of text into metadata.

It is time for the US courts to recognize explicitly that, in the digital age, copying books for non-expressive purposes is not infringement. Courts have already applied this logic in analogous cases: Google, Microsoft and others copy web pages to feed into their Internet search engines; the online service Turnitin copies exam papers and other sources so that plagiarism can be detected. These practices have been challenged and found to be legal under copyright law.

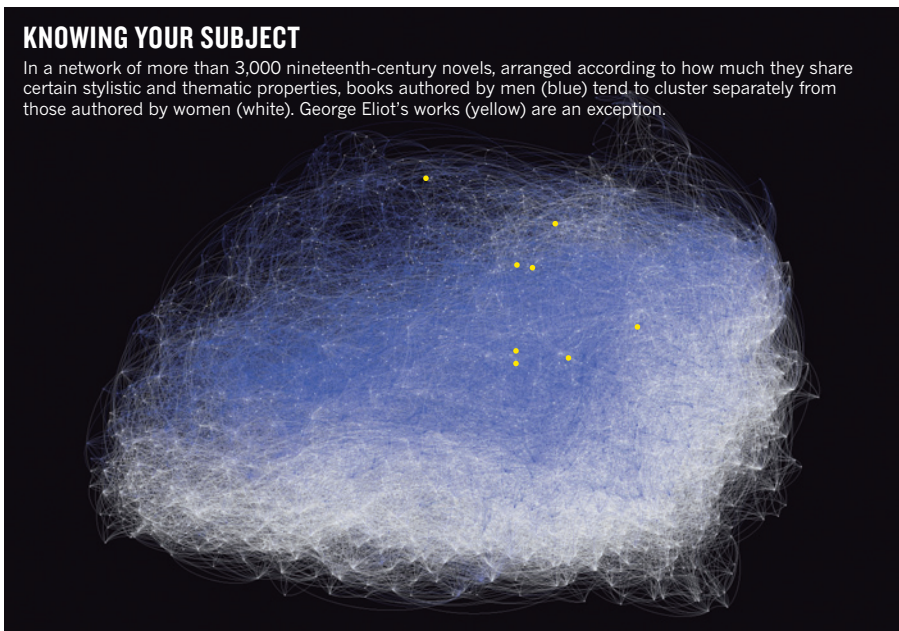
It is crucial for future research that the right precedent be set. We hope that the judges decide that digitization for text mining and other forms of computational analysis is, unequivocally, fair use. ■

Matthew L. Jockers is assistant professor of English at the University of Nebraska, Lincoln, USA. **Matthew Sag** is associate professor of law at Loyola University, Chicago, Illinois, USA. **Jason Schultz** is assistant clinical professor of law at the University of California, Berkeley, USA. e-mail: mjockers@unl.edu

1. Jockers, M. L., Sag, M. & Schultz, J. preprint at Social Science Research Network (2012); available at <http://ssrn.com/abstract=2102542>.
2. Hargreaves, I. *Digital Opportunity: A Review of Intellectual Property and Growth* (Intellectual Property Office, 2011).

KNOWING YOUR SUBJECT

In a network of more than 3,000 nineteenth-century novels, arranged according to how much they share certain stylistic and thematic properties, books authored by men (blue) tend to cluster separately from those authored by women (white). George Eliot's works (yellow) are an exception.



SOURCE: M. L. JOCKERS



Max Perutz, James Watson, John Kendrew and Francis Crick talk to a BBC presenter (centre) about their Nobel prizes in 1962.

What makes a great lab?

William Bynum reflects on the factors that have brought nine Nobel prizes to the UK Laboratory of Molecular Biology in Cambridge.

What makes an outstanding laboratory? There have been a number of these special places during the past couple of centuries, but none more so than the Laboratory of Molecular Biology (LMB) in Cambridge, UK, which is celebrating 50 years since it got its own building and, this week, 50 years since four of its scientists were awarded Nobel prizes — Max Perutz, John Kendrew, James Watson and Francis Crick.

Overall, the LMB can claim nine Nobel prizes for 13 staff scientists during its illustrious history, plus another eight for those who trained or worked there temporarily. As a unit within the Cavendish Laboratory at the University of Cambridge, the LMB had a distinguished pre-history. Few would have dared to predict that its independent existence would be so successful and productive. Seen historically, however, it shares characteristics with other outstanding laboratories. These include new methods of producing scientific knowledge, novel approaches to

training researchers, the innovation and excitement that surround an emerging scientific field and, perhaps most centrally, the presence of a gifted individual with the personality and vision to make things happen.

The LMB can thus be seen as a modern exemplar of a species that goes back to the early nineteenth century. Previous centres of excellence had differing ambiances, which reflected the scientific cultures of their times. But all share a few important star qualities. In search of these, here I briefly examine three of the LMB's ancestors and its parent laboratory.

LABORATORY LIFE

Before the nineteenth century, most laboratories were places in which single individuals worked, sometimes aided by an assistant or two. Chemist Antoine Lavoisier (1743–94), with his wife at his side, was typical. Wives and servants often helped. Justus von Liebig (1803–73) changed this pattern definitively.

Liebig's chemistry laboratory at the University of Giessen in Germany opened in

1826 and acquired international renown. It attracted students from all over Europe and earned Liebig a reputation as a 'chemist breeder'. His lab was an early example of the research and teaching establishments that made the German universities the envy of many. It began as a single room, with a fire in the middle surrounded by work benches. Liebig's work on the compositions of chemical substances and their reactions was outstanding, and his focus on agricultural, industrial and biological issues gave his research a highly topical flavour.

He trained his protégés carefully, especially in qualitative analysis. Students flocked to him, with the result that European chemistry during the middle of the nineteenth century bore a distinctly Liebigian flavour as his students moved to influential positions elsewhere. The identification of state-of-the-art problems and the training of students to solve them characterize his achievements. Liebig's initiative was widely adapted in the natural and biomedical sciences throughout the German university system.

Training was also part of the brief of physiologist Ivan Pavlov (1849–1936), but he brought his own organizational genius to bear on the 'physiology factory' that he masterminded in St Petersburg, Russia, famously studying dogs. He adapted aspects of manufacturing to the production of scientific knowledge. Pavlov's staff were among the first to specialize in different tasks: surgical, chemical, dog handling. The dogs were also specialized. Many had permanent gastric fistulas; some had oesophageal or pancreatic ones or other surgical interventions that allowed Pavlov to examine physiology *in situ*.

Pavlov's 'laboratory' eventually occupied a whole building — already nearer to the modern usage of the word, and a far cry from the single rooms of most researchers of earlier times. As science has become more complex and cooperative, so the physical structures of laboratories have evolved faster than the language we use to describe them.

The next example reinforces this point: Thomas Hunt Morgan (1866–1945) and his Fly Room at Columbia University in New York. It was more than just a room. Modern experimental genetics was born there, with the fruitfly (*Drosophila melanogaster*) as the prize experimental subject. The fly's rapid breeding time and four large chromosomes made it ideal for examining how chromosomal events during meiosis and mitosis relate to the structural features of the adult. A chance finding of a fly with a white eye — instead of the usual red — led Morgan to the importance of the sex chromosome.

Morgan was a gifted scientist who surrounded himself with equally gifted students and postdoctoral researchers, including Alfred Sturtevant, Calvin Bridges and Hermann J. Muller. Although Morgan was a

patrician from the US South, he ran his lab on egalitarian lines, with the consequence that historians still debate the relative contributions of the different parties. Morgan won a Nobel prize by himself in 1933, although he divided the money with Sturtevant and Bridges, to help to educate their children. Only Muller (who won his own Nobel in 1946 for his work on the effects of radiation on mutation rates) suggested seriously that Morgan sometimes exploited his students. Most believed that in the free interchange and mutual devotion to uncovering the genetics of the fruitfly was a formula that worked.

Liebig, Pavlov and Morgan each created something special. Their labs achieved international prominence, attracted talented scientists and bred further success. Each lab bore the stamp of the founder's ambitions and personality, and this relationship between the boss and the establishment stands out. Morgan's culture of egalitarianism provided a model for many successful modern laboratories, not least the LMB.

LAB SPECIATION

A kind of speciation can sometimes occur with laboratories. What is now the LMB began life ensconced in the Cavendish Laboratory, the centre of physics at the University of Cambridge and, by any reckoning, also among the most successful modern labs.

The Cavendish opened in 1874. Its first director, James Clerk Maxwell (1831–79), was arguably the most important physicist between Newton and Einstein. A genial man blessed with a fertile mind and remarkable ingenuity, Maxwell contributed to many problems in physics, and he completed the work on electromagnetism begun by Hans Christian Ørsted, Michael Faraday and others. He showed that the Sun's light comes to us through electromagnetic waves, and at the same time predicted the range of radiations that has been central to modern science and modern life.

Many of these advancements came from the Cavendish, beginning with J. J. Thomson (1856–1940) and his discovery of the electron in 1897. His was one of many Nobel prizes from the Cavendish, and like all leaders of successful laboratories, he was a good talent spotter. And successful labs attract ambitious and talented individuals. Ernest Rutherford (1871–1937) came from New Zealand to the Cavendish because of Thomson and his group. He thrived there, and after stints in Montreal, Canada, and Manchester, UK, he succeeded Thomson as director in 1919. He brought with him from Manchester James Chadwick, rooting nuclear and atomic physics firmly in the Cavendish during the early decades of the twentieth century.

Until it got its own building in 1962, the LMB was simply a research unit within the Cavendish. Of its early workers, the most

important was Max Perutz (1914–2002), whose style and personality shaped the LMB. Perutz came to England in 1936, hoping to work with Frederick Gowland Hopkins, the pioneer Cambridge biochemist. A meeting with the X-ray crystallographer J. D. Bernal, then still at the Cavendish, convinced the young Perutz that X-rays could provide the tools to solve the molecular structures of proteins. It took Perutz a further year to get horse haemoglobin crystals that were suitable for analysis using X-ray diffraction techniques. Because of the Second World War, it was seven more years before he could return to this molecule, his life's work. Even though he had lived in England for several years before the outbreak of war, he was treated as an enemy alien and incarcerated for nine months during 1940, in Britain and Canada. He spent the rest of the war back in Britain designing aircraft carriers.

Encouraged by Lawrence Bragg, then director of the Cavendish, Perutz returned to studying haemoglobin, joined by John Kendrew (1917–97). The beginnings of the LMB date from 1947, when the UK Medical Research Council (MRC) began supporting the work of Perutz and Kendrew. The original name of their group was the MRC Unit for Research on the Molecular Structure of Biological Systems. Perutz described himself then as a chemist working in a physics laboratory on a biological problem: a fairly accurate summary of the inputs into the field dubbed 'molecular biology' in 1938 by the Rockefeller Foundation administrator Warren Weaver.

Perutz and Kendrew pursued a promising avenue of molecular biological research, but haemoglobin is such a complex model that they soon added the simpler myoglobin to their agenda. Hugh Huxley joined

"A single administrator, Audrey Martin (and her dog Slippers), looked after things."

the group in 1948, but turned to studying the biophysical dynamics of muscle contraction: an early example of the widening range of biological problems in the unit's remit. The increasing international reputation of the group's work brought talented young scientists to Cambridge, including physicist Francis Crick as a research student and biologist James Watson as a postdoc.

A successful laboratory generally breeds more success, with implications for size and ethos. Bragg was appreciative of the group's prominence, but the Cavendish was chronically short of space in post-war austerity Britain. So in 1957 Perutz, aware that Frederick Sanger from the biochemistry department also needed more space, wrote to the MRC about housing the molecular biologists in a new laboratory. Unsurprisingly, negotiations were slow given so many vested interests, but

money was found. The present building, in Hills Road, Cambridge, expanded on more than one occasion and, still growing, was opened by Queen Elizabeth II in May 1962. There were then about 25 staff members and the same number of visiting workers. In October, there was a party to celebrate the Nobel prizes of Perutz and Kendrew in chemistry, and Watson and Crick (with Maurice Wilkins) in physiology or medicine.

BRIGHT BEGINNINGS

The party in 1962 was just the beginning. Over the past half-century the LMB has been at the centre of molecular biology, the discipline at the heart of the life sciences. New groups in developmental biology, immunology, cell biology and neurobiology attest to the expansion of the field. The growth has often been opportunistic, clustered around individuals such as Sydney Brenner, César Milstein, Aaron Klug and Michel Goedert. The simpler rules of only a generation ago (the original procedures for producing monoclonal antibodies were not patented, for instance) have given way to the contemporary competitive world of biotechnology.

As the laboratory has grown, its administrative structure has inevitably become more complex. Until Perutz retired in 1979, it had no director. Perutz didn't want to be one, and it meant he could retain his lab space after retirement. Instead, the lab had a loose management committee, which met occasionally and saw its main job as attracting outstanding talent to the lab. Perutz kept the bureaucratic structures of the laboratory minimalist, and until 1973 a single administrator, Audrey Martin (and her dog Slippers), looked after things. The egalitarianism that Morgan had fostered at Columbia was effectively duplicated in Cambridge, an ethos encouraged by the lab's successive directors — Sydney Brenner, Aaron Klug, Richard Henderson and Hugh Pelham — as each has presided over an ever-larger operation. The LMB now has some 400 workers, about half permanent staff and the rest students and visiting scientists.

It is easier to describe success than to explain it, but several of the characteristics that typified earlier exemplars are also part of the core ideology of the LMB. New techniques, new disciplines, new ways of tackling old problems and the ethos of collegiality have continued to characterize the lab. So has the key personality of Perutz, whose influence shines even after his death. The achievements of the lab's first half-century have been rewarded by even more new buildings, due to open in 2013. Long may the LMB flourish. ■

William Bynum is emeritus professor of the history of medicine at University College London, UK.
e-mail: w.bynum@ucl.ac.uk



Researchers in Côte d'Ivoire bag the skull of a colobus monkey to check for pathogens.

ZOOZOSIS

Fatal exchange

Nathan Wolfe applauds a tome on interspecies disease transmission that mixes research with human stories.

The exchange of microbes between humans and animals — zoonoses — began to fascinate me nearly 20 years ago, when I was studying wild primate populations. There are a remarkable number of zoonotic agents, ranging from anthrax to HIV, West Nile virus and influenza. It seemed shocking even in the early 1990s that this vastly important class of microbe was considered, if at all, to be a funny-sounding boutique niche in biology.

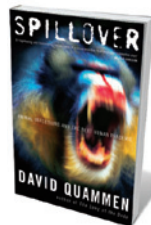
Fortunately, times have changed. Zoonoses have come of age in terms of scientific and public recognition. We now know that these agents are the most likely source of future pandemics. The past few years have seen an explosion in technical literature, government-sponsored research programmes, philanthropic interest and even films such as *Contagion*. During the past year, several general treatments of zoonoses have been published, including Jacques Pepin's *The Origin of AIDS* (Cambridge University Press, 2011), Craig Timberg and Daniel Halperin's *Tinderbox* (Penguin, 2012) and my own *The Viral Storm* (Allen Lane, 2011).

Now comes *Spillover* by David Quammen, one of that rare breed of science journalists who blend exploration with a talent for synthesis and storytelling. Quammen's excellent *The Song of the Dodo* (Prentice Hall & IBD, 1996), on island biogeography,

is difficult to top, but *Spillover* comes close. This is a timely, serious and impressive work that marks the maturation of a field of microbiology.

Quammen takes us into the field, offering an idea of research challenges in remote hotspots of disease emergence such as parts of Gabon and Malaysian Borneo. His narrative on the 'cut hunter' theory, based on our understanding of how HIV began, presents a realistic sense of the first person infected with the chimpanzee simian immunodeficiency virus (SIV) that would become HIV. It is the kind of portrayal that generalists and specialists have waited for.

In researching *Spillover*, Quammen visited scientists such as malaria researchers Janet Cox-Singh and Balbir Singh, and travelled to remote research sites in China, Bangladesh and beyond. He interviewed the usual suspects at leading international universities (myself included, in the context of pandemic prevention). But Quammen also sought out oft-neglected scientists and fieldworkers in places such as the Democratic Republic



Spillover: Animal Infections and the Next Human Pandemic
DAVID QUAMMEN
W. W. Norton: 2012
592 pp. \$28.95, £20

of Congo. These doughty professionals do considerable scientific heavy lifting, negotiating political and sometimes security hurdles to collect the samples that form the backbone of zoonosis research. Refreshingly, Quammen also highlights a number of up-and-coming young researchers in the field.

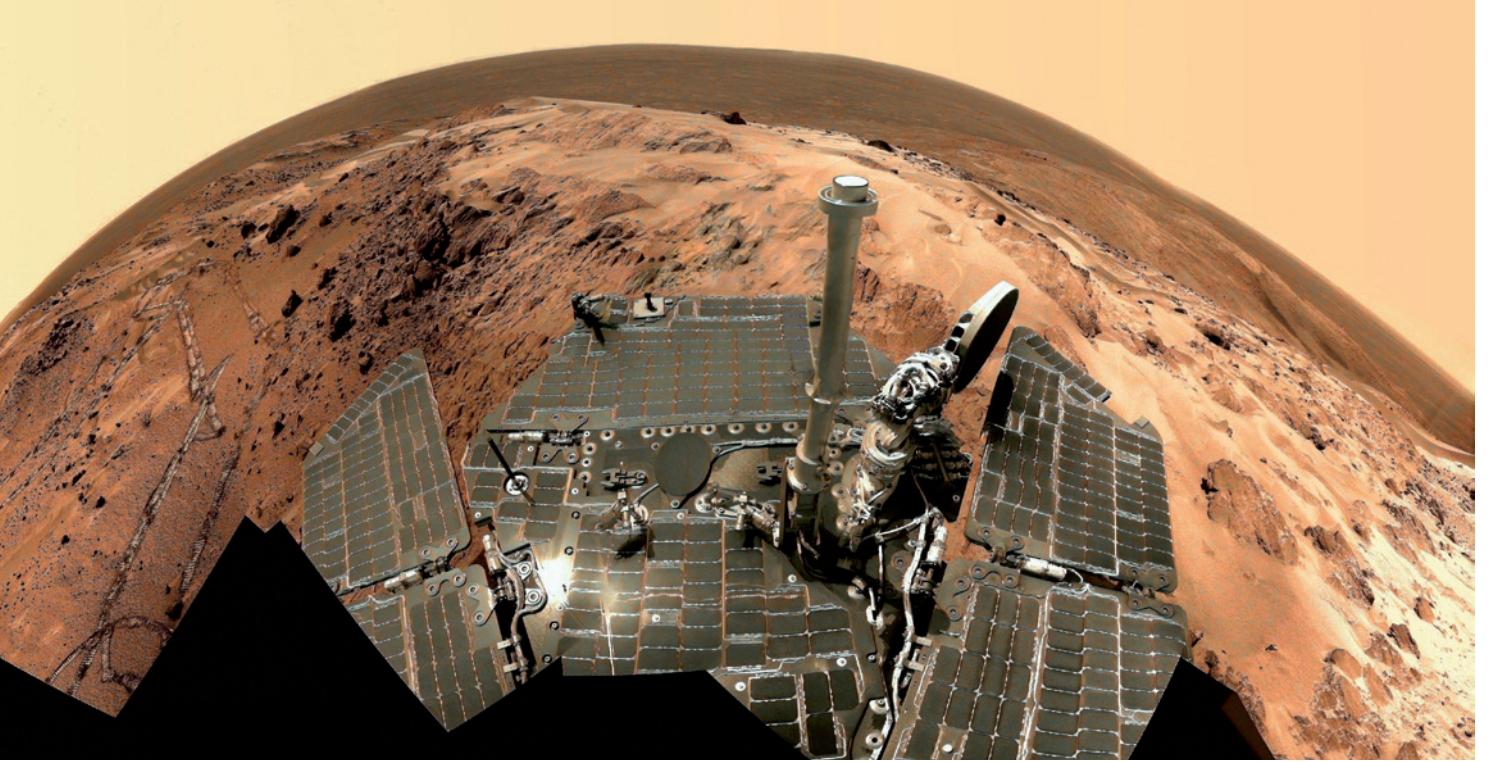
The book gives a dutiful review of the range of research on disease ecology and zoonosis, the subfields and their central concepts. But the behind-the-scenes stories are particularly enjoyable. Quammen brings to light the details left out of papers and technical talks; even workers in the field will find new stories. For example, his accidental encounter and subsequent interviews with survivors of the Mayibout Ebola epidemic in Gabon provide first-person perspectives on the epidemic and the hunting events that led up to it that are absent from any other account I've read.

Specialists will find little new in terms of scientific concepts. Nevertheless, some discussions, such as that on mathematical modelling in chapter three, push scientists towards the next level in their research. There are some minor technical errors, such as the age of the recombination event of monkey SIVs that produced SIV_{cpz}, the variant that infects chimpanzees; Quammen suggests that it may be only hundreds, rather than thousands, of years old. And some will bridle at a few oversimplifications — for example, the suggestion that the H5N1 flu virus is inherently more worrying than H1N1, which among other things belies the potential for reassortment between the two. But these are quibbles.

Importantly, *Spillover* challenges those working in the area of disease ecology and emerging infectious diseases to ask: what is next? Quammen concludes by reviewing some programmes that have been developed by bodies such as the US Department of Defense to understand and address the threats that zoonoses pose for the increasingly susceptible human population. These include the US Agency for International Development's Emerging Pandemic Threats Program, which aims to identify risks early and develop global capacity to stop them before they spread. Quammen rightly believes that such ventures are the way forward.

These efforts, and what they will evolve into, may be pivotal to human survival. As Quammen points out using the example of gypsy moths (*Lymantria dispar*) and nucleopolyhedroviruses, other species on have succumbed to pathogens following explosive population growth. A big question looms over all of us. How will humans fare? ■

Nathan Wolfe is a microbiologist at Stanford University in California. He is the founder of Global Viral and chief executive of Metabiota. His book *The Viral Storm* is shortlisted for the Royal Society's Winton Prize.
e-mail: ndwolfe@stanford.edu



NASA's Spirit rover landed in the Gusev Crater on Mars in January 2004.

MARS EXPLORATION

Roving the red planet

It is people who drive Curiosity and other robot missions on Mars, reminds **Jim Bell**.

There are Martians among us. Hundreds of them, living on Earth but working 'on' Mars, telecommuting to and fro as part of the interplanetary science and exploration programme conducted by NASA and other space agencies. These Martians are the men and women who help to control robots such as the Mars rovers on the red planet's surface, or any of the three satellites orbiting Mars and acquiring remote-sensing data. But both the media and the public tend to anthropomorphize the rovers as the explorers in this frontier world.

Cognitive scientist William Clancey, based at NASA's Ames Research Center in Moffett Field, California, works on how humans and societies adapt their lives and work to computers. During the early years of the Spirit and Opportunity missions, Clancey embedded himself in the Mars Exploration Rover (MER) science and engineering team at NASA's Jet Propulsion Laboratory in Pasadena, California. His *Working on Mars* documents firsthand many of the highs and lows of the people trying to carry out remotely controlled field science over a vast distance using mobile, programmable laboratories.

By observing these Martian "natives" and interviewing a subset of the science and engineering teams, Clancey captures some of the complex inner workings of a modern scientific expedition that just happens to be

on another planet. And he draws analogies to eighteenth- and nineteenth-century expeditions such as those of James Cook and Alexander von Humboldt.

The rovers' missions have become major media and sociological phenomena. The starry roster now includes the 1997 Mars Pathfinder Sojourner rover, the MER Spirit and Opportunity, which landed in 2004 (Opportunity is still active), and NASA's newest interplanetary vehicle, the car-sized



The team in charge of Curiosity celebrates the first pictures sent back from the red planet.

Curiosity rover that landed in early August. Millions of people worldwide, including educators and their students, have followed the missions online — an opportunity itself made possible mostly through the commitment of the rover science teams and NASA to share the adventures as widely as possible.

So the anthropomorphizing — which happens even among scientists working with the rovers — is perhaps not surprising. The robots are seen as having human-like senses: vision (cameras), touch (arms and drills) and taste (chemical analysis), as well as mobility (wheels). They are viewed as "plucky" or "intrepid", and during their missions "struggle" and "make discoveries". Among my colleagues on the rover science teams, it is common to refer to them as "robot geologists" or, in the case of Curiosity, a "robotic astrobiologist". Ultimately, even Clancey has to admit that casting the robots as valiant explorers in a dangerous land has heightened public interest in them — and perhaps even boosted public and Congressional support for NASA and its planetary-science missions.

Clancey underlines the importance of the unique and unprecedented sociotechnological fusion of remote sensors and the organizational structure, tactical-operations processes and strategic-planning tools pioneered by the MER team (many by trial and error and learning on the fly). Together, these promote the agency of the scientists and engineers and boost their capacity to actively and efficiently engage in scientific fieldwork in remote environments and situations.

One prime example is the way in which the MER planning software for rover driving and instrument deployment integrates the actual geology and

NATURE.COM
For more on the
Curiosity rover:
nature.com/curiosity

M. DI LORENZO ET AL./AVIATION WEEK/MARS EXPLORATION ROVER MISSION/CORNELL/JPL/NASA

B. VAN DER BRUG-POOL/GETTY



Working on Mars: Voyages of Scientific Discovery with the Mars Exploration Rovers

WILLIAM J. CLANCEY
MIT Press: 2012.
328 pp. \$29.95,
£20.95

topography of the rover's environment. Using previous days' imaging, it produces a fully functional computer-graphics representation of the vehicle in a virtual-reality environment; the result is almost like an interplanetary video game. The interaction between the virtual rover and its environment makes observation planning highly intuitive. It helps remote science team members to imagine being there in the field, which, in turn, helps to enable (for example, through guiding the rover's drives or the positioning of the instruments on specific parts of rocks or soils) many of the kinds of measurements that team members would actually make themselves if they were there.

This fusion provides a general model for successful future robotic, and human, remote field science. For example, the Curiosity mission's science and engineering team is using many of the operations and planning tools and structures of the MER mission, although necessarily modified for the specific attributes and constraints of that vehicle, landing site and mission objectives. Remote under-sea expeditions, as well as many other applications of telerobotics and telepresence, could similarly benefit from aspects of the MER model.

The human-centred computing aspects of missions such those of the Mars rovers have profound implications for the future of space exploration in general. For example, during the past few decades it has been popular to incite "humans versus robots" debates between advocates of crewed space-flight and of robotic exploration. As Clancey makes clear, the debate is moot.

Humans are already exploring the Solar System using the tools of robotics and the methods of field science. Robotic companions — "intrepid" or otherwise — are absolutely going to be part of humanity's eventual in-person exploration of our planetary neighbours. It is not going to be humans versus robots, but humans and robots, working together. ■

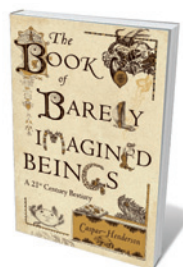
Jim Bell is a professor in the School of Earth and Space Exploration at Arizona State University in Tempe and a member of the camera teams for the *Sojourner*, *Spirit*, *Opportunity* and *Curiosity* Mars rover missions. His *Postcards from Mars* is a photographic history of the Mars Exploration Rover missions.
e-mail: jim.bell@asu.edu

Books in brief



Secrets of the Ice: Antarctica's Clues to Climate, the Universe, and the Limits of Life

Veronika Meduna YALE UNIVERSITY PRESS 232 pp. £29.95 (2012)
Bleak and storied wonderland it may be, but Antarctica is also the world's biggest lab, where hundreds of scientists cluster to take the planet's pulse. Science writer Veronika Meduna follows the action in this homage to frontier science. Beginning with geological and glaciological findings on Antarctica's climate history, she segues into marine life, the microbes of the cold deserts and the fugitive 'hum' of the Big Bang chased by physicists and astronomers on the continent. A beautifully illustrated journey through 'Mars on Earth'.



The Book of Barely Imagined Beings: A 21st Century Bestiary

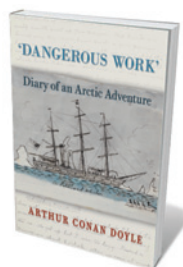
Caspar Henderson GRANTA 336 pp. £25 (2012)
Award-winning writer Caspar Henderson read Jorge Luis Borges's *The Book of Imaginary Beings* (1967) and realized that nature's creations often trump the fantastical for sheer surreality. Henderson's mainly marine beasts are a dazzling catch. The "genital fingered", gherkin-sized stomatopod *Gonodactylus smithii*, for instance, uses specialized limbs for defence — delivering enough force to break a bone. Eels, whales, arachnids and more are examined, with Henderson's central concern the survival of all this glory in the midst of the biodiversity drain. Wittily illustrated by Golbanou Moghaddas.



Guesstimation 2.0: Solving Today's Problems on the Back of a Napkin

Lawrence Weinstein PRINCETON UNIVERSITY PRESS 377 pp. £13.95 (2012)

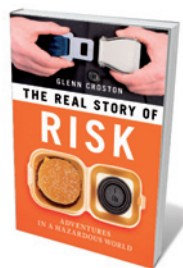
This follow-up to the popular *Guesstimation* offers more on the joy of mathematical estimation, and inspiration for the budding analyst. Physicist Lawrence Weinstein trawls questions from the pragmatic to the bizarre. Among them are his probings of energy, transportation and recycling such as gauging the US plastic-bag pile-up on the basis of hydrocarbon use. He also covers the senses, heavenly bodies, radiation — and the amount of urine in public swimming pools.



Dangerous Work: Diary of an Arctic Adventure

Arthur Conan Doyle (Eds Jon Lellenberg and Daniel Stashower)
BRITISH LIBRARY PUBLISHING 368 pp. £25 (2012)

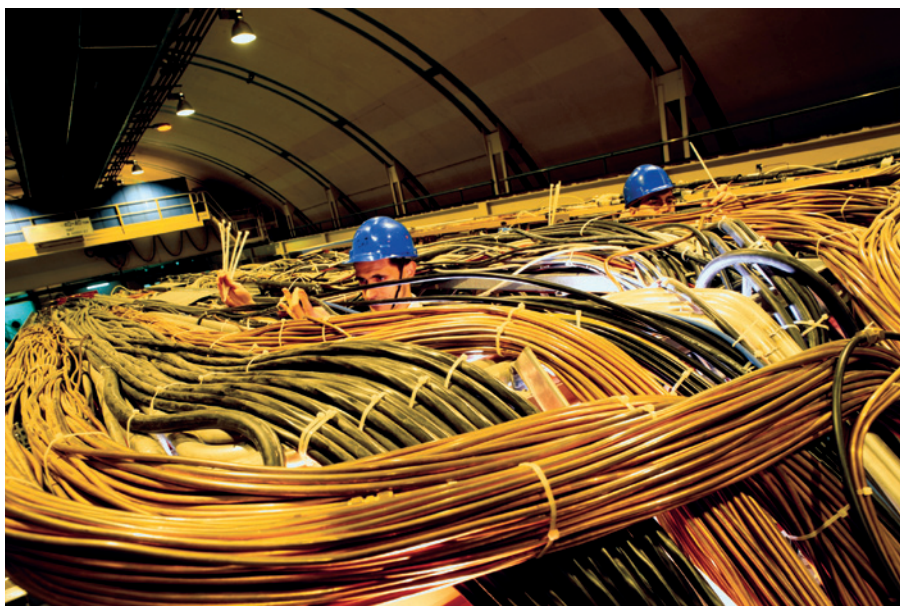
Who knew that Arctic explorers lauded the creator of fiction's most famous sleuth for his own detective work on routes to the North Pole? Arthur Conan Doyle — author of Sherlock Holmes — published the data in the article 'The Glamour of the Arctic' after a youthful stint as ship's surgeon on a Greenland whaler. His diary of the 1880 voyage is here reproduced in facsimile, with published pieces inspired by the trip. Hair-raising incidents abound, from a sudden on-board death by peritonitis to the young medic's periodic falls into ice-strewn waters.



The Real Story of Risk: Adventures in a Hazardous World

Glenn Croston PROMETHEUS BOOKS 256 pp. £16.99 (2012)

Imagine this: you find yourself worrying about shark attacks while crossing a busy road in a daze. Our perception of risk and the reality are frequently at odds, biologist Glenn Croston argues in this jazzyly written exploration of the balance between risk and reward. Croston marshals a raft of research on why our view of the phenomenon is so skewed, delving into evolutionary roots, our denial of 'slow' catastrophes, the role of lust in colouring our judgement, our need to belong and much more.



Supercomputers such as that at CERN in Switzerland are becoming faster at a predictable rate.

NETWORK THEORY

The regularities of facts

Carl Bergstrom assesses the power of scientometrics in predicting the shifts and shelf-life of knowledge.

How much of what you know today will be true tomorrow? In a week's time? In a decade? The weather forecast may change overnight; our estimate of the number of genes in the human genome may change in the coming months; our understanding of consciousness may be radically different a century from now.

Knowledge shifts over time, explains Sam Arbesman in *The Half-Life of Facts*, and it does so in predictable ways. The book takes us on a whirlwind tour of emerging fields of scientometrics, and undertakes a broader exploration of metaknowledge. Arbesman details how researchers beginning to focus the big-data lens back on science itself are uncovering quantitative laws and regularities in the way that scientific knowledge is constructed and modified over time.

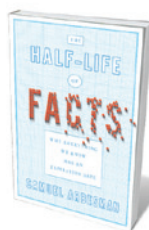
Like the decay of atoms, individual discoveries may be difficult to predict, but in the aggregate, facts change in highly regular ways. To illustrate this point, Arbesman ranges widely through the scope of human knowledge, drawing on examples from physics and chemistry, technology and medicine, sociology and cultural studies, and even the arts and humanities. For example, over time there is a predictable regularity to how measurement errors get smaller, how computation and travel become faster, how innovations diffuse through social networks

and how technological advances drive increases in human populations. Even the magnetic permeability of iron has increased in a consistent manner with changes in smelting technology.

Arbesman defines facts loosely, not as objective truths but as little pieces of knowledge, right or wrong. This casts a broad net over facts of many different kinds, at the risk of obscuring interesting and important distinctions.

Scientific 'facts' about the natural world — the nature of an electron or the evolutionary significance of a peacock's tail — change as science progresses and our explanatory frameworks shift. Statistics as facts — batting averages, gross national products, crime rates — change not because the body of knowledge around them changes, but because the world is changing beneath our feet and new events are transpiring.

So statistical facts need not be as interconnected as scientific facts; for example, Guinness World Records is a catalogue of



The Half-Life of Facts: Why Everything We Know Has an Expiration Date
SAMUEL ARBESMAN
Current: 2012.
256 pp. \$25.95,
£18.99

independent assertions. In 2011, Zac the macaw set the record for the most basketballs slam-dunked by a parrot in one minute; meanwhile, the world record for the fastest 100-metre hurdles wearing diving fins remained unchanged at 14.8 seconds.

But in proper network-theorist fashion, Arbesman focuses more on the similarities than on the differences between these distinct types of fact and processes of change. Both kinds are generated at predictable rates, change in predictable ways and are subject to scientometric analysis.

One quibble I have with the book is that occasionally I feel Arbesman's enthusiasm gets the better of him, and he accepts the conclusions of sound-bite science without adequate scrutiny. For example, if approximately 80% of citations are copied from earlier citations of the same material (M. V. Simkin and V. P. Roychowdhury *Complex Sys.* 14, 269–274; 2003), can we join Arbesman in the presumption that researchers read only 20% of what they cite? Or might this pattern arise because authors find it easier to compile their bibliographies from other reference lists — irrespective of whether they have read these papers?

Overall, however, Arbesman is a delightful guide to the territory, patently in love with this emerging field. He is also a skilled storyteller, and his wide-eyed reporting invigorates material that could have been dry and academic.

The chapter on hidden knowledge deserves particular note. It addresses one of the most pressing scientific problems we face: how to make vital new connections among ideas. In an era in which exhaustive reading is no longer possible and library-shelf browsing is infrequent, how can we design mechanisms that connect scholars with the ideas that they need to move forwards?

Arbesman hints at possible elements of a solution: innovation prizes, social tagging, systematic meta-analyses and automated discovery programs. Big changes are coming in the very near future, driven by the confluence of the digital revolution in publishing, the explosion of computational capacity and the accumulated strain of a 350-year-old system of scientific communication pushed to the breaking point in an exponentially larger world.

The current generation will solve these problems, and change how science as a social and communicative process is practised. Reading *The Half-Life of Facts*, I became excited about the prospect of living through — and perhaps even contributing to — this change. ■

Carl Bergstrom is a network theorist and professor of evolutionary biology at the University of Washington in Seattle.
e-mail: cbergst@u.washington.edu

Correspondence

Biodiversity needs a scientific approach

We agree with Esther Turnhout and colleagues (*Nature* **488**, 454–455; 2012) that the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) should take citizen knowledge and non-monetary values into account to improve the science-policy interface for biodiversity protection. Even so, knowledge used to inform policy must be produced through an objective process if it is to withstand scrutiny. This demands a science-based approach.

Science sets a standard for data quality, not for who collects the data. It provides a common currency for understanding the consequences for biodiversity of actions arising from the values of different stakeholders, including local communities, hunter-gatherers, commercial exploiters and conservationists.

The role of IPBES in policy formulation means that it will inevitably meet resistance that will seek to undermine data credibility, the assessment process and the platform itself. Instead of avoidance strategies, we need mechanisms for successful negotiation of such controversies to support transformation.

David A. Westcott, Frederieke J. Kroon, Andy W. Sheppard
CSIRO Ecosystem Sciences, Australia.
david.westcott@csiro.au

Change of heart on nanoparticle risks

You contend that most nanotechnology researchers now acknowledge that some areas of their work raise legitimate environmental, health and safety concerns (*Nature* **488**, 576–579; 2012). This was not the case a decade ago, when we at the Action Group on Erosion, Technology and Concentration (ETC) called for a moratorium on the commercialization of

products containing engineered nanoparticles.

In 2002, scientists could point us to only one peer-reviewed study of nanotube toxicity, and companies were still sending a Material Safety Data Sheet for graphite with carbon nanotube shipments. ETC's concerns were dismissed as alarmist. We welcome the change in attitude.

ETC's central concern has always been the economic impact on populations in developing countries resulting from the market disruptions that are expected with the advent of new nanoproducts and processes. We have consistently dismissed the hypothetical concept of 'grey goo' — uncontrolled self-replicating nanorobots — as a red herring.

Finally, ETC has no connection to ITS, the group that claimed responsibility for the nanotech-related bombings in Mexico. ETC opposes violence in all forms.

Silvia Ribeiro *ETC Group, Mexico City, Mexico.*
etc@etcgroup.org

Fleischmann denied due credit

Philip Ball's obituary of Martin Fleischmann (*Nature* **489**, 34; 2012), like many others, ignores the experimental evidence contradicting the view that cold fusion is 'pathological science' (see www.lenr.org). I gave an alternative perspective in my obituary of Fleischmann in *The Guardian* (see go.nature.com/rzuckfz), describing what I believe to be the true nature of what Ball calls a "Shakespearean tragedy".

The situation at the time of the announcement of cold fusion was confused because of errors in the nuclear measurements (neither Fleischmann nor his co-worker Stanley Pons had expertise in this area) and because of the difficulty researchers had with replication. Such problems are not unusual in materials science. Some were able, I contend, to

get the experiment to work (for example, M. C. H. McKubre *et al.* *J. Electroanal. Chem.* **368**, 55–56; 1994; E. Storms and C. L. Talcott *Fusion Technol.* **17**, 680; 1990) and, in my view, to confirm both excess heat and nuclear products.

Scepticism also arose because the amount of nuclear radiation observed was very low compared with that expected from the claimed levels of excess heat. But it could be argued that the experiment never excluded the possibility that the liberated energy might be taken up directly by the metal lattice within which the hydrogen molecules were absorbed.

In my opinion, none of this would have mattered had journal editors not responded to this scepticism, or to emotive condemnation of the experimenters, by setting an unusually high bar for publication of papers on cold fusion. This meant that most scientists were denied a view of the accumulating positive evidence.

The result? Fleischmann was effectively denied the credit due to him, and doomed to become the tragic figure in Ball's account. **Brian D. Josephson** *University of Cambridge, UK.*
bdj10@cam.ac.uk

European biodiesel can be sustainable

An accurate evaluation of the sustainability of European oilseed rape for biodiesel production would be a useful resource in discussions of the European Union's bioenergy policies. Your ill-judged pronouncement in an online News report that rapeseed biodiesel fails the sustainability test (*Nature* <http://doi.org/jdn>; 2012) risks confusing the facts by quoting questionable figures from a preliminary study (G. Pehnelt and C. Vietze *Jena Econ. Res. Pap.* 039; 2012).

These figures concern some of the most important parameters used in sustainability calculations. The study

considerably underestimated mean annual seed yields of rapeseed used for biodiesel, by using outdated yield values from the entire European Union (around 2.8 tonnes per hectare for 1991–2005), rather than current yields from the principal biodiesel-producing countries such as Germany (3.8 tonnes per hectare for 2005–10; see <http://faostat3.fao.org>). The input values were also based on energy-intensive production procedures (deodorization, for example) that are only used in processing rapeseed oil for food, and on unrealistic transportation emission values. Incorrect input data can seriously influence the outcome of a sustainability evaluation.

Political decisions need to be based on reasoned and constructive discussion about issues as controversial as renewable biofuels, which in turn must be based on strong, peer-reviewed science.

Rod Snowdon, Wolfgang Friedt
Justus Liebig University, Giessen, Germany.
rod.snowdon@agrar.uni-giessen.de

Reviews turn facts into understanding

Your Editorial on *h*-index forecasting (*Nature* **489**, 177; 2012) perpetuates the myth that: "Review articles, which may not add much to the research, count the same as original research papers, which contribute a great deal." Probably the reverse is true. A research paper usually provides just one or two new facts, whereas reviews synthesize our understanding more broadly and make it more concrete.

Some reviews summarize thousands of papers (see, for instance, D. B. Kell *BMC Med. Genom.* **2**, 2; 2009) and turn an inchoate and stochastic scientific literature into knowledge.

Douglas B. Kell *University of Manchester, UK.*
dbk@manchester.ac.uk

FORUM: Climate science

The aerosol effect

Anthropogenic aerosols in the atmosphere undoubtedly influence climate. But do the approaches taken in climate models to account for the effects of aerosols provide meaningful estimates of those effects? Two climate scientists offer their opinions.

THE TOPIC IN BRIEF

- Aerosol particles in the atmosphere influence clouds, and thereby climate, because they act as nuclei for cloud formation.
- Computational models of climate systems have sought to incorporate the effects of aerosols on clouds through parameterizations.

- However, the representation of aerosol–cloud interactions in climate models is based on simplifications that ignore the complexity of the small-scale physical processes governing such interactions in the real world.
- The value of studying the effects of aerosol–cloud interactions in climate models has therefore been questioned.

better or more reliable, just more complex. Additional complexity can be great fun, but it should not disguise the fact that, at least for aerosol–cloud interactions, much is speculative, and the results of such complex models should be taken with a grain of salt.

Bjorn Stevens is at the Max Planck Institute for Meteorology, Hamburg 20146, Germany. e-mail: bjorn.stevens@mpimet.mpg.de

Grains of salt

BJORN STEVENS

There is something captivating about the idea that fine particulate matter, suspended almost invisibly in the atmosphere, holds the key to some of the greatest mysteries of climate science. Recent studies have reported that interactions of such aerosols with clouds may be hiding a large part of the sensitivity of global temperature to increasing levels of greenhouse gases¹. It has also been claimed that aerosol–cloud interactions are reshaping patterns of rainfall² and even influencing the development of tornadoes³. But such interactions invariably turn out to be more nuanced than the simple ideas underpinning these and related studies would lead one to suspect. This explains why, despite decades of research, no consensus has emerged as to exactly what, other than perhaps a slight mitigation of twentieth-century global warming, is attributable to aerosol–cloud interactions⁴.

To put matters into context, consider the analogous question of the effect of carbon dioxide on climate. Carbon dioxide always has the same composition, its lifetime is very much longer than that of atmospheric circulation systems, and it affects fluxes of radiant energy in ways that are well understood and are not particularly contingent on other factors. By contrast, aerosol particles can differ widely in composition, are ephemeral and affect clouds (and hence radiative fluxes) in ways that are poorly understood and depend on a long list of factors.

To determine the correct sign — let alone the magnitude — of the effect of some important aerosol–cloud interactions, one may, to borrow words from elsewhere, need a weatherman to know which way the wind blows⁵. Models can make good weathermen, and thus provide information about the wind, as well as about many of the other factors on which aerosol–cloud interactions depend. But for climate models, this is true only on scales that are so large (hundreds of kilometres) as to be almost irrelevant. On fine scales (tens of metres), at which there is some understanding of how aerosol particles influence cloud microstructure, climate modellers are groping in the dark.

Aerosol–cloud interactions are not only contingent on, but also compound, the ‘cloud question’. It is one thing to ask, as the cloud question does, how the distribution of clouds responds to robust changes in the large-scale environment — for instance, changes that follow temperature. It is quite another to ask how indeterminate changes in clouds will be modified by uncertain interactions with aerosol particles whose properties are also not well understood. So, although some answers to the cloud question seem to be within reach, a quantitative understanding of the global effect of aerosols on clouds is definitely not. Fortunately, given that global aerosol burdens have remained more or less constant for a decade or longer⁶, and that greenhouse-gas concentrations are increasing relentlessly, answering the cloud question might help to answer the more pressing questions related to Earth’s changing climate.

Our poor understanding of the global effect of aerosols on clouds means that the incorporation of additional details into climate models does not make the models fundamentally

An essential pursuit

OLIVIER BOUCHER

Aerosols and clouds influence each other through multiple interactions at temporal and spatial scales that span many orders of magnitude. Increasingly detailed observations and modelling studies have revealed the complexity of aerosol–cloud interactions. Such complexity was certainly not recognized when the ‘climate forcing’ associated with these interactions — that is, their effects on the amount of energy entering and escaping from the atmosphere as radiation — was first estimated in an atmospheric general circulation model⁷. After almost two decades of research, parameterizations of these interactions in large-scale models remain simplistic in some respects, but they are neither worthless nor useless.

Estimates of anthropogenic climate forcing are essential for constraining climate sensitivity to such forcing in observations⁸. They are also important for understanding recent climate change⁹ and to improve regional decadal climate prediction. Although small-scale observations and models are invaluable for investigating how man-made aerosols might affect cloud properties, they cannot easily be scaled up because of the large variability of aerosols, clouds (Fig. 1) and environmental conditions, and because they overlook the feedbacks on aerosols and clouds that act through large-scale atmospheric dynamics.

Satellite observations have been widely used to infer correlations between aerosols



Figure 1 | Cloud diversity. Climate models must represent the net effect of clouds on solar and thermal irradiances over scales that often encompass a rich variety of cloud types, such as the diverse range in the cloud field pictured. Aerosol particles affect the microphysical structure of clouds, as well as irradiances in the cloud environment. The net effect of these interactions on the radiative properties of clouds is poorly understood, and incorporating them realistically into climate models is a challenge.

and cloud properties or precipitation. They can be combined with observations of Earth's radiative budget — the amount of energy that enters and leaves Earth's atmosphere as radiation — to estimate aerosol forcing on a global scale. The methodological challenges of such an approach, however, are increasingly being recognized, because both aerosols and cloud properties depend hugely on meteorological conditions. Climate models, although imperfect, are therefore indispensable for calculating climate forcing by anthropogenic aerosols. Such models are the only means of representing the feedbacks induced by aerosol–cloud interactions at the global scale. They can also be used to tease apart the interwoven effects of weather and aerosols on clouds, which currently impede observational studies¹⁰.

Estimates of climate forcing from aerosol–cloud interactions have not varied much over time, even though research has moved from empirical to more mechanistic approaches for parameterizing these interactions. This does not mean, however, that such estimates are robust. They often lack a proper treatment of uncertainties, and there are indications from temperature observations that many climate-model simulations of such interactions imply too much cooling.

It is therefore vital to continue with efforts to parameterize aerosol–cloud interactions in climate models as a means of overcoming the present limitations. Indeed, some promising avenues of research are opening up. It is now possible to perform high-resolution simulations that incorporate complex microphysical representations of aerosol–cloud interactions over sufficiently large areas, and for long enough periods, to investigate the couplings

between a cloud field and its environment, thus going some way towards bridging the gap between the cloud scale and the global scale. New methods for representing the statistics of cloud properties at scales below those resolved by climate models have also been developed^{11,12}. Finally, cloud microphysics

is now recognized as being important for numerical weather prediction; systematic verification of weather-model outputs against observations, and powerful data-assimilation techniques, may provide new insight into aerosol–cloud interactions. With so many encouraging developments, let's embrace the challenge! ■

Olivier Boucher is at the *Laboratoire de Météorologie Dynamique, IPSL/CNRS, Université Pierre et Marie Curie, 75252 Paris Cedex 05, France.*
e-mail: olivier.boucher@lmd.jussieu.fr

- Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T. & Bellouin, N. *Nature* **484**, 228–232 (2012).
- Koren, I., Feingold, G. & Remer, L. A. *Atmos. Chem. Phys.* **10**, 8855–8872 (2010).
- Rosenfeld, D. & Bell, T. L. *J. Geophys. Res. Atmos.* **116**, D20211 (2011).
- Levin, Z. & Cotton, W. R. (eds) *Aerosol Pollution Impact on Precipitation: A Scientific Review* (Springer, 2009).
- Khain, A. *Environ. Res. Lett.* **4**, 015004 (2009).
- Zhang, J. & Reid, J. S. *Atmos. Chem. Phys.* **10**, 10949–10963 (2010).
- Jones, A., Roberts, D. L. & Slingo, A. *Nature* **370**, 450–453 (1994).
- Schwartz, S. E. *Surv. Geophys.* **33**, 745–777 (2012).
- Kaufmann, R. K., Kauppi, H., Mann, M. L. & Stock, J. H. *Proc. Natl Acad. Sci. USA* **108**, 11790–11793 (2011).
- Quaas, J., Stevens, B., Stier, P. & Lohmann, U. *Atmos. Chem. Phys.* **10**, 6129–6135 (2010).
- Wang, M. *et al.* *Atmos. Chem. Phys.* **11**, 5431–5455 (2011).
- Rio, C. & Hourdin, F. *J. Atmos. Sci.* **65**, 407–425 (2008).

MUCOSAL IMMUNOLOGY

Infection induces friendly fire

Our immune system usually ignores 'friendly' gut bacteria. But when infection with a pathogen damages the intestine's mucosal lining, the resident microbes can invade the body, inducing immune responses directed at themselves.

DAVID MASOPIST & VAIVA VEZYS

The mammalian gastrointestinal tract harbours an extensive microbial community that is usually well tolerated by the immune system. Sometimes, however, inappropriate immune responses directed against these 'commensal' organisms arise, such as in inflammatory bowel disease. But how such responses are triggered is unclear. Writing in *Science*, Hand *et al.*¹ demonstrate that pathogenic infections that disrupt the integrity of the gut's mucosal barriers can precipitate the development of long-lived immunity directed against the host's resident microbes.

CD4⁺ T cells are a class of immune cell that

deploys potent effector functions to combat infections. During development, a repertoire of T cells is generated such that there will be low numbers of cells specific for any given antigen — a substance that stimulates responses from T and B cells of the immune system. 'Naïve' T cells are those that have not yet encountered the antigen for which their own antigen receptor is specific. Infection with a pathogen stimulates proliferation of pathogen-specific T cells, which not only increases the number of these cells, but also causes their differentiation to effector cells. As the proliferative response is ending, a population of long-lived pathogen-specific 'memory' cells is established, which provides a heightened state

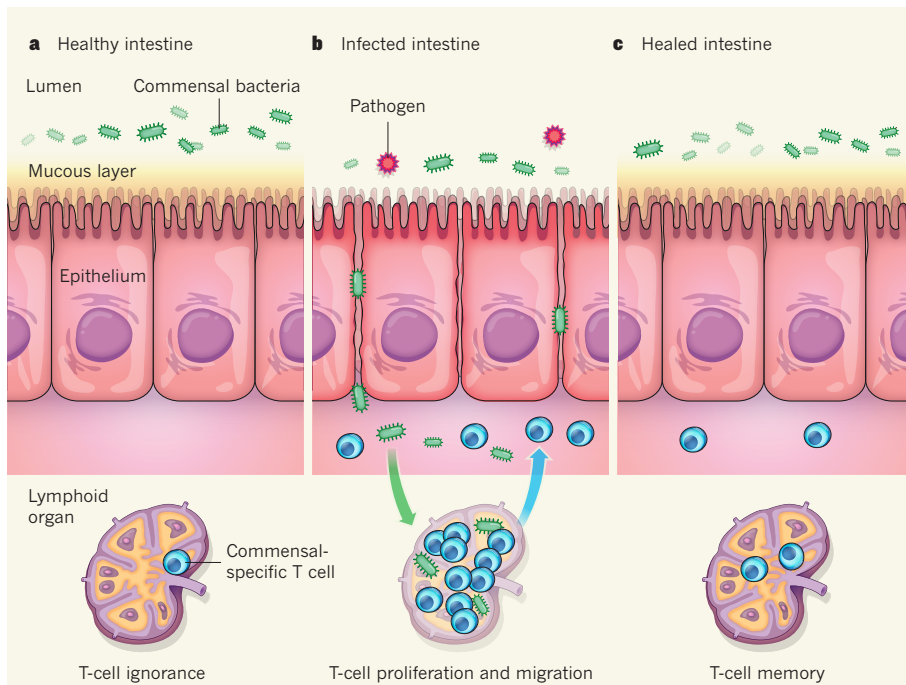


Figure 1 | Breaching the barrier. **a**, The lumen of the mammalian intestine contains abundant non-pathogenic bacteria, which are called commensal bacteria. Although there are CD4⁺ T cells in distant lymphoid organs (such as the spleen or lymph nodes) that specifically recognize these 'friendly' bacteria, these T cells typically ignore the bacteria and remain in a naive state. **b**, Hand *et al.*¹ show that infecting mice with a protozoan pathogen that causes intestinal damage allows the commensal bacteria to migrate to other parts of the body. This results in activation and proliferation of both pathogen-specific (not shown) and commensal-specific CD4⁺ T cells in the lymphoid organs, and migration of these cells to the intestine. **c**, The authors also show that a population of commensal-specific memory T cells persists in the intestine long after the resolution of infection and restoration of the mucosal barrier.

of readiness to respond to repeat infections. The signals received from different microbes cause the responding CD4⁺ T cells to adopt effector functions best suited to controlling the infection². However, if a naive T cell encounters its specific antigen in the absence of infection, it is rendered 'tolerant', preventing a misdirected immune response. Thus, the context of initial antigen exposure is pivotal in determining T-cell behaviour.

Hand *et al.* examined CD4⁺ T cells specific for the flagellin protein expressed by certain *Clostridium* bacteria that are commensals in the gut lumen of adult mice (Fig. 1). They found that these non-pathogenic organisms are typically ignored by T cells, which remain in their naive state. However, when the researchers infected the mice with the parasitic protozoan *Toxoplasma gondii*, which causes intestinal damage, the clostridia breached the mucosal barrier and could be recovered from lymphoid organs such as lymph nodes and spleen. This translocation of the commensal bacteria caused naive *Clostridium*-flagellin-specific CD4⁺ T cells residing in lymphoid organs to proliferate and migrate to the gut.

Interestingly, as the *Clostridium*-specific T cells proliferated, they differentiated into effector T cells in a manner that mimicked the CD4⁺ T-cell response to *T. gondii*. They upregulated expression of T-bet, a gene-regulatory

protein, and they acquired the ability to produce the cell-signalling molecule interferon- γ . These behaviours are characteristic of the Th1 differentiation state, which is associated with immune responses to intracellular infections but has also been implicated in numerous autoimmune diseases. Hand and colleagues also show that the mice retained a population of flagellin-specific memory CD4⁺ T cells long after resolution of the *T. gondii* infection and re-formation of the mucosal barrier.

These results have several interesting ramifications. For instance, inflammatory bowel disease (IBD), which includes Crohn's disease and ulcerative colitis, is characterized by aberrant CD4⁺ T-cell responses directed against intestinal commensal bacteria³. Although it is not clear how and why these responses are initiated, it is known that variations in several genes, including ones that regulate the immune response to gut bacteria and maintenance of the mucosal barrier, are tightly linked to IBD incidence⁴. However, identical twins with equal genetic susceptibility to IBD do not always both experience disease, which suggests a critical role for environmental triggers in this condition⁵.

Hand *et al.* provide evidence that exposure to pathogenic microbes could provide such a trigger. First, their results demonstrate that CD4⁺ T-cell ignorance of commensal bacteria

can be broken during an infection, probably because of the increased amount of microbial antigen that is available to stimulate the immune system. Second, they show that pathogenic infections shape the differentiation program of contemporaneous commensal-specific CD4⁺ T-cell responses. In other words, when the commensal-specific CD4⁺ T cells are first activated, it is in the context of infection, and so they adopt an effector program that is dictated by the offending pathogen rather than by the resident gut microbes.

This differentiation program is distinct from that seen in experimental situations in which CD4⁺ T cells are exposed to commensal microbes in the absence of pathogenic infection^{6–9}. It will be interesting to examine the relationship between various infections and the development of IBD-like disease in genetically predisposed mouse models. This is likely to be a fruitful line of enquiry, because a recent study has shown that infection of genetically susceptible mice with murine norovirus tipped the balance towards Crohn's-like disease symptoms¹⁰.

Hand and colleagues' study also emphasizes the fact that memory CD4⁺ T cells specific for commensal microbes may exist. The memory T-cell population studied by the authors, which was specific for a single peptide from *Clostridium*, was fairly small. However, when one considers that more than 3 million non-redundant genes have been identified¹¹ in the human intestinal microbiome (all the microorganisms in the gut), this raises the question of whether a substantial fraction of all memory CD4⁺ T cells may actually be directed against commensal microbes. If this is the case, it would be particularly interesting to examine the effect of these cells on immune responses to subsequent intestinal infections (either with the same pathogen or one not previously encountered) that also cause disruptions of the mucosal barrier. It is conceivable that the commensal-specific memory CD4⁺ T cells have a protective capacity, by amplifying immunological sensing of barrier disruption. However, a commensal-reactive immune T-cell population could also contribute to inflammatory or autoimmune diseases of tissues exposed to such microbes, which include the skin and upper respiratory tract.

Thus, in addition to illuminating a potential mechanism underlying inflammatory conditions such as IBD, Hand and colleagues' findings indicate a potential limitation of the 'specific-pathogen-free' mouse model in which most basic immunological experiments are conducted. Such mice are not typically exposed to gut infections that compromise barrier integrity. This point provocatively suggests that, when studying immune responses to pathogens or commensal organisms, immunologists should also consider examining mice that have experienced past intestinal infections, because such infections could have a lasting

impact on the immune system and might better recapitulate the human condition. ■

David Masopust and Vaiva Vezys are in the Department of Microbiology, Center for Immunology, University of Minnesota, Minneapolis, Minnesota 55455, USA.

e-mails: masopust@umn.edu; vvezys@umn.edu

1. Hand, T. W. *et al.* *Science* **337**, 1553–1556 (2012).
2. Pepper, M. & Jenkins, M. K. *Nature Immunol.* **12**, 467–471 (2011).
3. Duchmann, R. *et al.* *Clin. Exp. Immunol.* **102**, 448–455 (1995).
4. Khor, B., Gardet, A. & Xavier, R. J. *Nature*

- 474**, 307–317 (2011).
5. Spehlmann, M. E. *et al.* *Inflamm. Bowel Dis.* **14**, 968–976 (2008).
6. Lathrop, S. K. *et al.* *Nature* **478**, 250–254 (2011).
7. Atarashi, K. *et al.* *Science* **331**, 337–341 (2011).
8. Geuking, M. B. *et al.* *Immunity* **34**, 794–806 (2011).
9. Ivanov, I. I. *et al.* *Cell* **139**, 485–498 (2009).
10. Cadwell, K. *et al.* *Cell* **141**, 1135–1145 (2010).
11. Qin, J. *et al.* *Nature* **464**, 59–65 (2010).

QUANTUM PHYSICS

Cruise control for a qubit

Continuous feedback control — monitoring a system and adjusting its dynamics — is widely used to keep systems ‘on track’. This approach has now been used to maintain the cycling of a quantum bit almost indefinitely. SEE LETTER P.77

HOWARD M. WISEMAN

Imagine you have to test-drive a car on a smallish circular racetrack, making exactly one circuit every minute. There is a big clock on a tower by the start/finish line. What strategy do you adopt? An obvious one is to keep comparing your position around the track with the position of the second hand around the clock. If your circular position lags that of the second hand, then step on the accelerator; if it leads, then ease off (Fig. 1). This illustrates three typical features of a process known as continuous feedback control: monitoring, comparison with a goal and dynamical adjustment¹. On page 77 of this issue, Vijay *et al.*² apply this type of control to the most basic quantum system: a two-state system called a quantum bit (qubit). By monitoring the excitation of their qubit with quite high efficiency, inferring its lag or lead relative to a radio-frequency field (their clock), and modulating the power applied to their qubit, they maintain a regular cycling between the qubit's excited and ground states almost indefinitely.

In an ideal world, when the power supplied by a car's engine equals the power dissipated (for example, by air resistance and friction), the car's speed will remain steady. But in the real world, unpredictable fluctuations in the car and its environment (such as variations in wind speed, air pressure, tyre pressure, oil temperature, road surface and other variables) will cause the speed to vary even with constant power. With no feedback control, the position of the

car around the track would, within an hour, become completely unrelated to the position of the second hand on the reference clock. In most modern cars, cruise control can do a good job of keeping the speed within a narrow range around a set value. But for the task in hand, a more sophisticated cruise-control mechanism would be needed, which compares

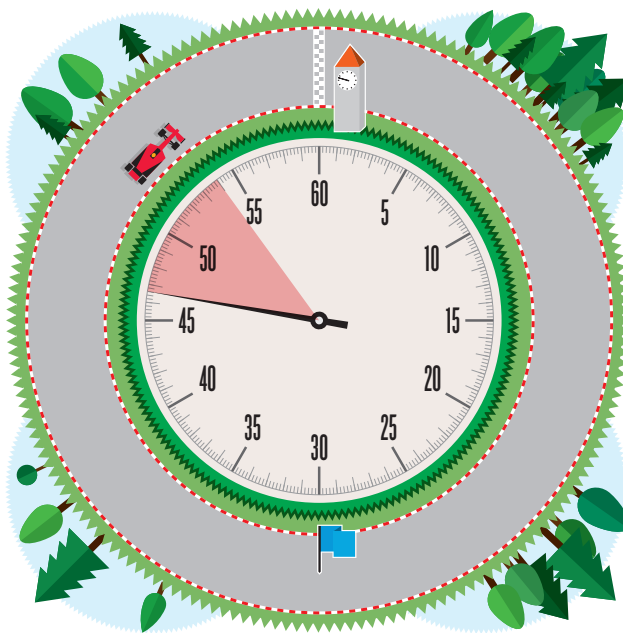


Figure 1 | Locking cycles to an external clock. Vijay and colleagues² used continuous feedback control to keep a quantum bit (qubit) cycling between its ground and excited state. The process is analogous to a car driver trying to make one lap every minute on a circular racetrack by monitoring the position of the car on the track and comparing it with the position of the second hand of a clock on a tower by the start/finish line. In this illustration, the car is slightly ahead of where it should be, so the driver should ease off the accelerator. In this manner, the driver can continuously compensate for any fluctuations affecting the car's speed. In a system that dissipates power, such as the qubit, fluctuations are inevitable, and are also introduced by the measurement of the qubit's state. In this analogy, the qubit's excited state is equivalent to the car being at a point (blue flag) that is diametrically opposite to the start/finish line, which corresponds to the ground state.

the position of the car with the position of the second hand, as described above.

In their experiment, Vijay *et al.*² use a transmon³, a tiny superconducting device that can be restricted (at ultra-low temperatures) to just two quantum states: its ground state and its first excited state. These two states define the qubit. The qubit dynamics in this experiment is in many ways analogous to that of the car on the circular track. If the start/finish line is like the ground state of the qubit, the point diametrically opposite is the excited state. The other points around the track are analogous to superpositions of the excited and ground states — they are still definite ('pure') states of the qubit, but, unlike the ground and excited states, they do not have a definite energy value.

By applying a microwave driving field, the qubit can be made to cycle around the 'superposition track' at a particular frequency, clockwise from ground state to excited state and back again. But like a car, the qubit also suffers from power dissipation. In the quantum world, this necessarily gives rise to fluctuations, even at zero temperature¹. As a consequence, after several microseconds the state of the qubit on its cycle will be completely unpredictable.

It is not obvious a priori that using continuous measurement and feedback control to lock the qubit cycling to that of an external clock will work, because in the quantum regime monitoring induces fluctuations too¹. However, as the original theoretical proposal for this quantum feedback protocol showed⁴, the process should work well for strong and efficient monitoring.

To monitor their transmon qubit, Vijay *et al.* coupled the qubit to a microwave probe field in a superconducting resonator that has a higher frequency than the driving field. The probe field continuously leaks out of the resonator, and is amplified into a macroscopic current. The authors used a simple feedback-control algorithm to modulate the driving field in proportion to the deviation of the observed current from an ideal sinusoidal current generated by the radio-frequency field acting as the external clock.

So how well does it work? Although

the monitoring was strong (compared with the dissipation), its efficiency was less than 50%. As a result, the feedback control could not do better than to keep the qubit state more or less in the correct half (as defined by the external clock) of the superposition track at any given time. Moreover, for about 10% of the time the transmon was found in a higher excited state than its first excited state, outside the qubit's two states altogether. But compared with the no-feedback result of complete unpredictability within several microseconds, the observed stabilization of the qubit's cycling is a big step forward in the feedback control of an individual qubit.

This work paves the way to many more

experiments on qubit feedback measurement and control, which would be enabled by more efficient monitoring than that achieved here. These experiments include rapid purification of a qubit⁵, tests of quantum-jump theory⁶, feedback control based on measurement back-action⁷ and feedback control based on the quantum Zeno effect⁸. In the longer term, continuous feedback control of multi-qubit systems is one path to correcting quantum errors in a quantum computer. With Vijay and colleagues' experiment, solid-state physics joins quantum optics at the forefront of quantum feedback-control investigations. ■

Howard M. Wiseman is at the Centre for

Quantum Computation and Communication Technology (Australian Research Council), and the Centre for Quantum Dynamics, Griffith University, Brisbane 4111, Australia. e-mail: h.wiseman@griffith.edu.au

1. Wiseman, H. M. & Milburn, G. J. *Quantum Measurement and Control* (Cambridge Univ. Press, 2010).
2. Vijay, R. *et al.* *Nature* **490**, 77–80 (2012).
3. Koch, J. *et al.* *Phys. Rev. A* **76**, 042319 (2007).
4. Ruskov, R. & Korotkov, A. N. *Phys. Rev. B* **66**, 041401(R) (2002).
5. Jacobs, K. *Phys. Rev. A* **67**, 030301(R) (2003).
6. Wiseman, H. M. & Gambetta, J. M. *Phys. Rev. Lett.* **108**, 220402 (2012).
7. Jacobs, K. *New J. Phys.* **12**, 043005 (2010).
8. Tanaka, S. & Yamamoto, N. preprint at arxiv.org/abs/1208.4197 (2012).

of hits. To reduce the complexity of this gene set and to remove redundant information, Qin *et al.* created metagenomic linkage groups (MLGs) of genes that showed similar profiles in either the patient or the control samples, in terms of taxonomic assignment or relative abundance of the markers (Fig. 1). This approach addressed two major problems with current metagenomic analyses. First, analyses so far have depended on comparisons with available microbial reference sequences to assess the species and functional composition of the metagenome. Such studies are limited by the fact that many detected sequences do not map to these references, so uncharacterized or rare microbes may go undetected. But by using MLGs, rare gene regions can be associated with a larger MLG, and therefore taken into account. Second, associations based on taxonomy alone can be misleading, as gene transfer between species or variation in gene content within a species can complicate taxonomic assignments.

One of the associations with type 2 diabetes identified by Qin and colleagues was a depletion of butyrate-producing bacteria. Butyrate is a molecule used as an energy source by cells in the gut lining, but it has also been suggested to reduce inflammatory responses in the colon⁵. Consequently, a decrease in butyrate may contribute to the increased intestinal inflammation implicated in insulin resistance⁵ and type 2 diabetes. Qin *et al.* also saw an increase in colonization by opportunistic pathogens and an enrichment in microbial genes related to oxidative-stress resistance in people with the disease; both of these changes may further contribute to the inflammatory environment.

However, before the results of this study can be accepted as providing generalizable risk factors for type 2 diabetes, further investigations will be necessary. First, obesity was not a factor in this study, which — given the link between type 2 diabetes and obesity — limits its applicability. Similarly, diet, which is known to have a strong influence on the composition of the gut microbiota⁶, must be considered in

GENOMICS

Resident risks

An innovative method for probing the genomes of the vast community of microorganisms that inhabit the human gut provides an alternative approach to identifying risk factors for type 2 diabetes. SEE LETTER P.55

JULIA OH & JULIA A. SEGRE

The interplay between genetics and the environment in controlling our health remains an open area of investigation. Technological advances in genome sequencing and analysis have empowered the search for the genetic determinants of human traits and diseases, but genetic variants that directly cause disease have so far been identified for only a small proportion of common diseases. For many other conditions, variants that increase disease risk are known, but the incompleteness of these associations suggests that environmental factors significantly contribute to the likelihood of development of these diseases.

A candidate source of environmental influence on disease development that is gaining increasing attention is the human microbiome — the trillions of microorganisms that inhabit our bodies. Our primary genome is in continual interaction with the combined genomes of these residents, which makes this 'metagenome' a fascinating part-environmental, part-genetic factor in human biology. On page 55 of this issue, Qin *et al.*¹ present an innovative method for searching the microbial metagenome for disease risk factors, using as their example type 2 diabetes, a disease that is reaching epidemic proportions in many regions across the world.

Heritability estimates for the common forms of type 2 diabetes range between 30 and 60%, suggesting that it is a complex disorder with multiple interacting genetic and environmental components. Genome-wide association

studies (GWAS), which are used to investigate individuals' DNA sequences for genetic variants associated with disease, have identified variants that contribute to the development of type 2 diabetes, but these only partially explain the variation in individual risk^{2,3}. Other significant risk factors for the disease, such as obesity and diet, are also associated with an altered gut microbiome⁴, suggesting that this disease is a good candidate in which to search for risk factors associated with the human microbial metagenome.

Qin *et al.* surveyed the gut metagenomes (derived from stool samples) of 345 Han Chinese individuals with and without type 2 diabetes using a method called shotgun sequencing, which produces random fragments of DNA sequence from the complex mix of microbial genomes. These fragments are then assembled into larger, gene-length fragments on the basis of sequence overlap. The authors compared these larger fragments with reference genome sequences and gene-function databases to determine the proportions of different bacterial species present in the samples, and consequently the relative representation of various microbial functional and metabolic attributes. They then used these 'markers' in a similar way to how variations at specific nucleotides are used in GWAS to identify blocks of DNA sequences, and carried out a 'metagenome-wide association study' (MGWAS) to identify different features of the metagenomes in people with and without the disease.

The MGWAS identified more than 42,000 gene markers associated with type 2 diabetes — a massive and unwieldy number

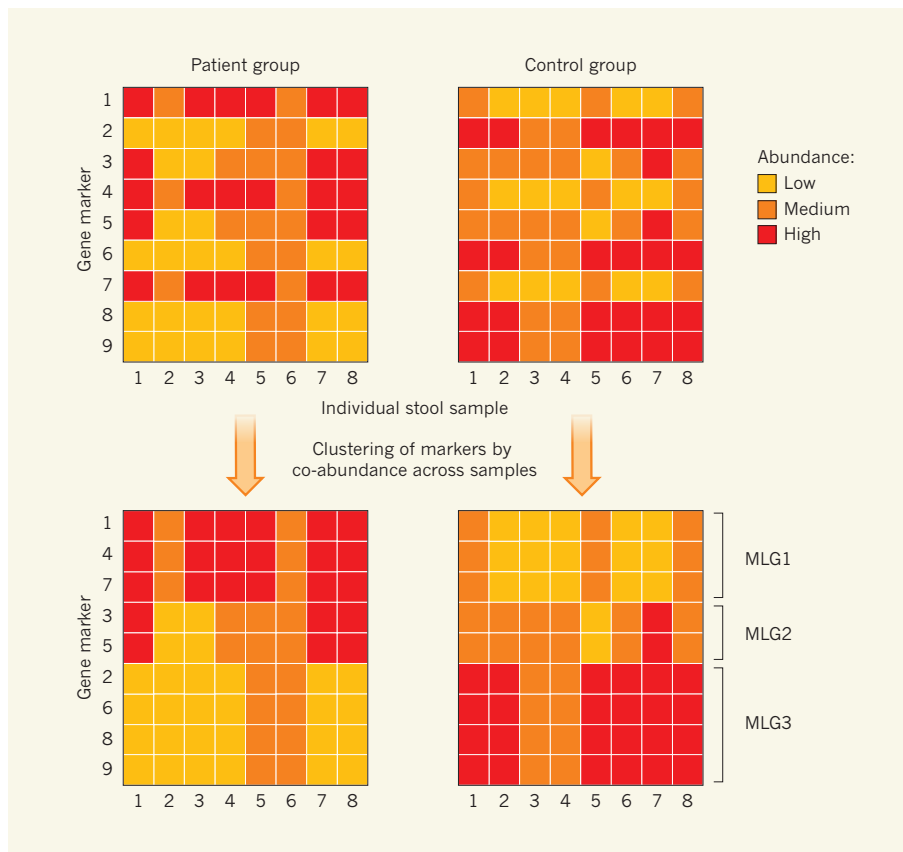


Figure 1 | Mining the metagenome. Qin *et al.*¹ studied the metagenome of the human gut — the combined genomes of the trillions of its resident microorganisms — in the search for microbial species and gene functions that are associated with type 2 diabetes. First, they identified microbial gene markers and determined the abundance of these markers in stool samples from a group of patients with type 2 diabetes and from a control group of people without the disease. The authors then clustered these markers according to their relative abundance, and assigned markers that had similar abundance profiles within either the patient or the control group to a metagenomic linkage group (MLG). In this example, MLG1 is over-represented in the patient group, MLG2 is similarly represented in patient and control groups and MLG3 is over-represented in the control group. The use of MLGs reduced the number of associations for subsequent analysis, minimized redundancy and allowed rare gene markers to be taken into account.

future research. Butyrate is a by-product of the metabolism of plant-based foods, but this raises a ‘chicken-and-egg’ question of whether individuals with type 2 diabetes have a diet lacking in these foods to begin with, leading to differences in the composition of their gut microbiomes, or whether it is the microbiota that contributes to disease development. To clarify some of these questions, it would be of great interest for future investigations to cross ethnic lines and to follow patients on controlled diets over extended periods. Furthermore, it is important to note that this study was associative rather than causal, and that controlled prospective studies are needed to better distinguish a metagenomic contribution from potentially confounding factors such as diet and obesity.

Beyond its contribution to our understanding of type 2 diabetes, this study introduced a novel approach — the application of traditional elements of GWAS to metagenomic data, and the introduction of MLGs to analyse the association of metagenomic variants with

a disease. This method will be applicable to a wide range of diseases or human traits that are likely to be influenced by the microbiome. The potential interactions between human genetic variation and metagenomic variation are staggeringly complex, and we are only beginning to understand how the human genotype influences the characteristics of its resident microbiota, and how, in turn, the microbiome can influence human health.

Finally, although MGWASs have great potential to account for some of the ‘missing heritability’⁷ in diseases such as type 2 diabetes, this same concept also leads to questions about whether the microbiome, and its effect on human traits, is itself heritable. Infants inherit, or converge on, similar microbiota to that of their parents, through breast milk, physical contact and shared living. But the gut microbiome is also amazingly plastic, responding rapidly to such factors as dietary changes and antibiotic treatment. And so our growing understanding of the metagenome may in some cases further blur, rather than

define, the boundaries between genetic and environmental influence on human traits and disease. ■

Julia Oh and Julia A. Segre are at the National Human Genome Research

Institute, National Institutes of Health, Bethesda, Maryland 20892-4442, USA. e-mails: julia.oh@nih.gov; jsegre@nhgri.nih.gov

1. Qin, J. *et al.* *Nature* **490**, 55–60 (2012).
2. Morris, A. P. *et al.* *Nature Genet.* **44**, 981–990 (2012).
3. Frayling, T. M. *Nature Rev. Genet.* **8**, 657–662 (2007).

4. Turnbaugh, P. J. *et al.* *Nature* **457**, 480–484 (2009).
5. Roelofs, H., Priebe, M. & Vonk, R. *Benefic. Microbes* **1**, 433–437 (2010).
6. Turnbaugh, P. J. *et al.* *Sci. Transl. Med.* **1**, 6ra14 (2009).
7. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. *Proc. Natl Acad. Sci. USA* **109**, 1193–1198 (2012).

ASTROPHYSICS

Two black holes found in a star cluster

The detection of two candidate black holes in a dense system of stars in the Milky Way suggests that a larger population of such objects might be lurking in this system. [SEE LETTER P.71](#)

STEFAN UMBREIT

The Milky Way globular cluster M22, a massive, dense star cluster containing close to a million stars, is starting to reveal its secrets. On page 71 of this issue, Strader *et al.*¹ present long-exposure radio images of M22, obtained by the Karl G. Jansky Very Large Array of radio antennas, showing two previously unknown sources of radio waves. The sources are candidates for ‘stellar-mass’ black holes, with masses 10–20 times that of the Sun. These are the first possible black holes to be detected in massive, old Milky Way star clusters such as M22. Furthermore, by virtue of the fact that two black holes were found, and were found in the cluster centre, the discovery provides insight into the dynamic evolution of the black-hole population.

Although the existence of black holes was predicted by theory almost a century ago², it was only in the past three decades that observational evidence became available. Because black holes are dark, the only way to detect them is through their gravitational influence on surrounding matter. In most cases, this requires that the black hole is in a binary system with another object that is either transferring mass to the black hole (accreting), at the same time emitting a large amount of energy as X-rays, or that it is bright enough to allow measurements of its radial velocity to be made.

In the present study, however, the main evidence for the interpretation of the two new sources as black holes comes from the relationship between the observed

radio emission and the X-ray emission. The radio emission from a black hole is usually attributed to radiation from jets of gas emanating from either side of a disk of gas, which is accreted from a companion star (Fig. 1). The X-ray emission results from strong shear in the inner part of the disk and from gas turbulence, which together heat the gas to such high temperatures that it releases X-rays.

Although many of the details of the connection between the gas jet and disk remain unclear, observations³ of accreting, stellar-mass black holes led to the determination of a relationship between the X-ray and radio

emissions, for low accretion rates. Thus, as X-ray luminosity decreases, the radio emission becomes increasingly dominant. Perhaps most importantly, observations⁴ as well as theoretical studies⁵ show that the radio-to-X-ray emission ratio increases with black-hole mass. This makes radio observations ideally suited to detect stellar-mass black holes⁶.

Strader *et al.* argue that the fact that the two new radio sources were not detected by the Chandra X-ray satellite places an upper limit on the sources’ X-ray luminosity. When combined with their radio emission, this limit yields a large minimum radio-to-X-ray ratio, corresponding to stellar-mass black holes each with 10–20 solar masses. The interpretation of the sources as two black holes is compelling, not least because further support comes from, among other independent evidence, their location close to the cluster centre. In a ‘self-gravitating’ stellar system (in which all the individual components are held under the combined gravity of the object as a whole) which is in thermal equilibrium, the average distance of an object from the centre is a function of its mass, with more-massive objects lying farther in. Given that the core of M22 would have reached approximate thermal equilibrium in a relatively short time (0.3 billion years) compared with the cluster’s age (12 billion years), the positions of the two sources can be used to determine the masses of the black holes. Using this approach, the authors deduced that the black holes are about 15 times more massive than the Sun — which is in accord with their previous calculation of 10–20 solar masses.

The detection of two accreting black holes in a globular cluster has implications for our understanding of the structure and dynamic evolution of dense stellar systems. First, there may be more than two black holes in M22, either as single entities or in binary systems in which mass accretion does not occur. Theoretical calculations⁷ of the formation of binary systems composed of a black hole and a white-dwarf star in globular clusters indicate that, over a period of 10 billion years, only 2–40% of the total retained black-hole population forms black-hole–white-dwarf

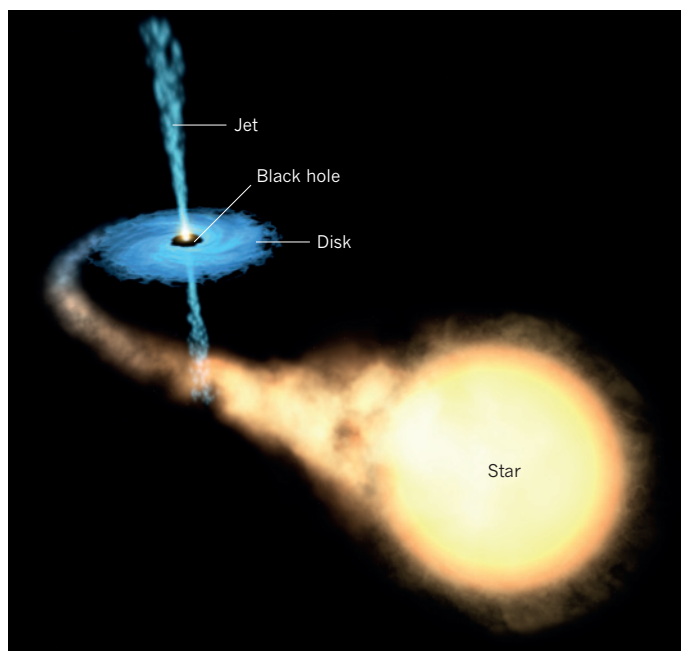


Figure 1 | Artist's impression of an accreting black hole. Strader *et al.*¹ detected two stellar-mass black-hole candidates in a Milky Way globular cluster by analysing the systems’ radio and X-ray emissions. The radio emission is associated with a jet that emanates from both sides of a disk of gas, which is attracted to the black hole from a companion star. The X-ray emission is linked to the inner part of the disk.

F. MIRABEL (CEA & INST. ASTRON. SPACE PHYS.)/CONICET ARGENTINA/ESA/NASA

binaries that have observable gas accretion. Therefore, if the sources detected in M22 were in binaries with white dwarfs, M22 could contain as many as 100 black holes. Second, simulations of clusters that have a large population of stellar-mass black holes showed⁸ that the black-hole population leads to a substantial expansion of the cluster core. This is mainly caused by frequent ejections of the black holes from the core into the outer cluster regions by close gravitational encounters with other black holes. Strader *et al.* suggest that this expansion could explain why M22 has the fifth-largest core radius among bright Milky Way globular clusters.

Finally, and perhaps most significantly, Strader *et al.* point out that the discovery of two stellar-mass black holes in M22 challenges a hypothesis that has been held for decades — that the population of black holes disappears rapidly through gravitational interactions, such that only one, or a binary system of two, black holes remain at the typical age of globular clusters (10^{10} years)^{9–11}. This would be all the more important if the black holes detected in M22 were part of a large population; Strader and colleagues' results indicate that more black holes may be coupled to the cluster.

If many black holes are retained in globular clusters, we would expect an increased detection of gravitational waves from the merging of black-hole binaries. Although the interaction rate between black holes, and consequently the formation of black-hole–black-hole binaries, would be larger if the black holes were

decoupled from the cluster, the rate of destruction of potential black-hole binaries would likewise increase, leading overall to fewer merger events¹². Future gravitational-wave searches will allow this expectation to be tested, and the full impact of Strader and colleagues' findings to be revealed. ■

Stefan Umbreit is at the Center for Interdisciplinary Exploration and Research in Astrophysics, and in the Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA.
e-mail: s-umbreit@northwestern.edu

1. Strader, J., Chomiuk, L., Maccarone, T. J., Miller-Jones, J. C. A. & Seth, A. C. *Nature* **490**, 71–73 (2012).
2. Schwarzschild, K. Sber. *K. Preuss. Akad. Wiss.* **7**, 189–196 (1916).
3. Gallo, E., Fender, R. P. & Pooley, G. G. *Mon. Not. R. Astron. Soc.* **344**, 60–72 (2003).
4. Merloni, A., Heinz, S. & Di Matteo, T. *Mon. Not. R. Astron. Soc.* **345**, 1057–1076 (2003).
5. Heinz, S. & Sunyaev, R. A. *Mon. Not. R. Astron. Soc.* **343**, L59–L64 (2003).
6. Maccarone, T. J. *Mon. Not. R. Astron. Soc.* **360**, L30–L34 (2005).
7. Ivanova, N. *et al.* *Astrophys. J.* **717**, 948–957 (2010).
8. Mackey, A. D., Wilkinson, M. I., Davies, M. B. & Gilmore, G. F. *Mon. Not. R. Astron. Soc.* **386**, 65–95 (2008).
9. Kulkarni, S. R., Hut, P. & McMillan, S. *Nature* **364**, 421–423 (1993).
10. Sigurdsson, S. & Hernquist, L. *Nature* **364**, 423–425 (1993).
11. Kalogera, V., King, A. R. & Rasio, F. A. *Astrophys. J.* **601**, L171–L174 (2004).
12. Downing, J. M. B., Benacquista, M. J., Giersz, M. & Spurzem, R. *Mon. Not. R. Astron. Soc.* **416**, 133–147 (2011).

IMMUNOLOGY

Tolerating pregnancy

The activity of specific suppressive immune cells, some of which persist to aid subsequent pregnancies, helps to explain how a pregnant female's immune system tolerates fetal antigens inherited from the father. SEE LETTER P.102

ALEXANDER G. BETZ

Pregnancy poses a conundrum for the immune systems of placental mammals. A pregnant female's immune system has to defend both mother and fetus from pathogens, while at the same time tolerating the fetus, which contains antigens that the maternal immune system recognizes as foreign because they are the products of genes inherited from the father. On page 102 of this issue, Rowe *et al.*¹ demonstrate that, during pregnancy, immune cells called regulatory T cells that recognize these paternal antigens proliferate in the mother and specifically suppress the maternal immune response against the fetus. Furthermore, the authors show that a pool of these cells remains long after delivery, facilitat-

ing tolerance in subsequent pregnancies. The description of this mechanism may, in the future, help to develop treatments for pre-eclampsia and prevent miscarriages resulting from immune rejection of the fetus².

Genetically, a fetus is half mother, half father. From an evolutionary perspective, maternal exposure to paternal antigens in the fetus is a relatively new problem: most animals lay eggs, so tolerance is not an issue. Yet physical attachment of the developing mammalian fetus to the mother's uterine wall by the placenta provides clear benefits — it allows gas exchange, nutrient uptake and waste disposal through the mother's blood circulation, providing optimal conditions for the growth of the developing fetus. A systemic immune suppression to facilitate this fetal 'implantation' would be much too



50 Years Ago

La Vie par Jean Rostand et Andrée Téry — Anyone glancing through this comprehensive, abundantly illustrated and well-produced volume could be forgiven for thinking life was very odd. Scarcely a page is passed that does not contain a picture of the abnormal, from the slightly unusual to the chimerical ... It begins with morality and religion, passes to the consideration of the living cell and the latest knowledge on this important topic ... and ends with some speculations on the future of the human race. One can but envy the authors their erudition — or the amount they must have learned in compiling the volume — and congratulate them on the fruits of their labours.

From *Nature* 6 October 1962

100 Years Ago

Satisfied with the maintenance of a specious standard of chemical purity, the public has acquiesced in the elevation of sky-scrapers and the sinking cavernous places of business. Many have thus become cave-dwellers, confined for most of their waking and sleeping hours in windless places, artificially lit, monotonously warmed. The sun is cut off by the shadow of tall buildings and by smoke — the sun, the energiser of the world, the giver of all things which bring joy to the heart of man, the fitting object of worship of our forefathers ... Modern civilisation has withdrawn many of us from the struggle with the rigours of nature ... I maintain that the bracing effect of cold is of supreme importance to health and happiness, that we become soft and flabby and less resistant to the attacks of infecting bacteria in the winter, not because of the cold, but because of our excessive precautions to preserve ourselves from cold.

From *Nature* 3 October 1912

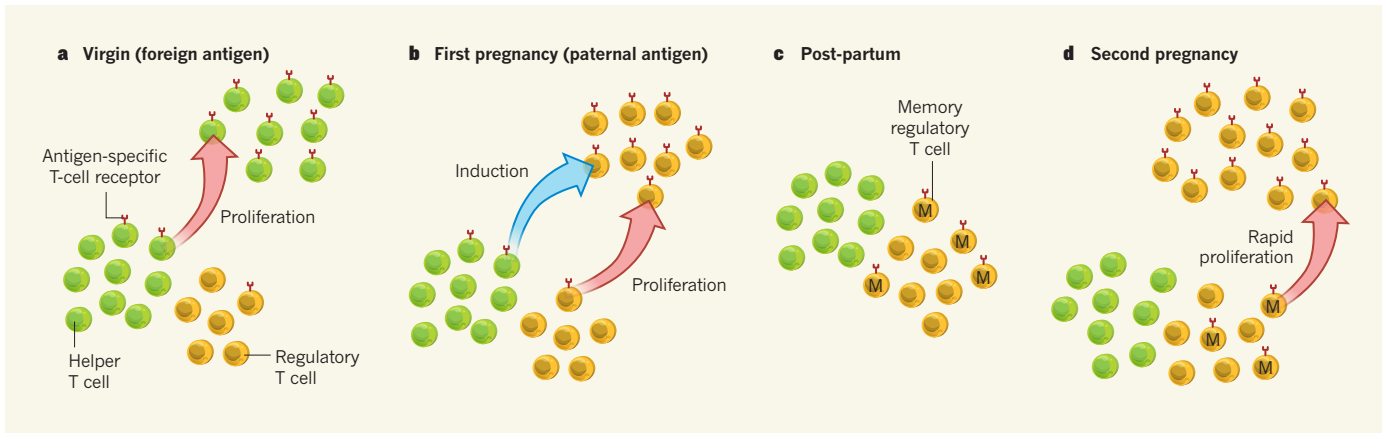


Figure 1 | Immune memory for a father's antigens. Rowe and colleagues¹ show that the mouse immune response to a foreign antigen differs radically depending on the context. **a**, In virgin mice, antigen-specific helper T cells proliferate in response to infection with *Listeria* bacteria expressing the antigen, despite the presence of a small number of regulatory T cells with matching antigen specificity (regulatory T cells are suppressive immune cells that dampen the responses of other immune cells). **b**, By contrast, when the same antigen is presented by the fetus during

pregnancy, antigen-specific regulatory T cells accumulate, as a result of both proliferation of existing regulatory T cells and the differentiation of regulatory T cells from helper T cells (induced regulatory T cells). **c, d**, Rowe and colleagues also show that a memory pool of these antigen-specific regulatory T cells persists long after delivery (**c**) and that, in subsequent pregnancies with the same father, these memory regulatory T cells rapidly proliferate and provide immune tolerance to the same fetal antigen (**d**).

risky because it would expose the mother and developing offspring to infection. So placental animals had to evolve a mechanism for localised and specific immune suppression.

Rowe and colleagues have helped to clarify this mechanism. It was previously known that the mother's immune system tolerates the fetus despite being fully aware of it³ and that a class of immune cell called regulatory T cells has a key role in this process⁴. Regulatory T cells act as suppressors of immune responses⁵; they are best known for their role in the prevention of autoimmune responses, but they are also involved in other functions, such as shutting down immune responses after successful elimination of a pathogen. These cells differentiate from precursor T cells in response to expression of just a single protein, Foxp3 (ref. 5), either in the thymus during T-cell development (thymic regulatory T cells) or in peripheral immune organs, such as the spleen or lymph nodes, during immune responses (induced regulatory T cells).

To assess how regulatory T cells promote immune tolerance of a fetus, Rowe *et al.* studied the fate of antigen-specific T cells responding to the antigen when encountering it either as a pathogen antigen or as a paternally derived antigen during pregnancy. When the authors infected mice with *Listeria* bacteria expressing the antigen, they observed antigen-specific 'helper' T cells proliferating — the expected immune response to a foreign antigen. But the immune response they saw in pregnant mice when the same antigen was expressed by the fetus was characterized by a substantial expansion in the number of maternal antigen-specific regulatory T cells (Fig. 1). This was the result of both proliferation of antigen-specific thymic regulatory T cells, and the conversion of antigen-specific helper T cells in peripheral

organs to regulatory T cells through the induction of Foxp3. The authors show that this immune suppression is highly antigen-specific, which explains why the pregnant female's ability to launch immune responses against infections is not affected.

These results demonstrate that the context of pregnancy can determine whether a foreign antigen is attacked or tolerated. The crucial role of regulatory T cells in this process provokes the question of whether these cells might have influenced the evolution of placental implantation of the fetus. Two recent comparative-genomics studies^{6,7} provide some clues. Although a Foxp3-like gene is present in fish, the protein it encodes lacks key domains required for the commitment of T cells to the regulatory lineage and for their function⁶. Intriguingly, it seems that Foxp3 was lost from the genome of birds, but retained in mammals, in which it acquired additional functional domains⁶ and additional elements that regulate its expression⁷. Foxp3 in monotremes (egg-laying mammals) contains most, if not all, of the functional domains⁶, which indicates that regulatory T cells evolved before the evolution of fetal implantation. Thus, it is tempting to speculate that the evolutionary gain of these cells facilitated the evolution of invasive placentation, because it provided the maternal immune system with a mechanism to tolerate the fetus without unduly compromising its responsiveness to pathogens.

Intriguingly, Rowe *et al.* also show that the number of fetal-antigen-specific regulatory T cells persists at elevated levels long after delivery of the baby, and that the cells rapidly proliferate during subsequent pregnancies (Fig. 1). This is reminiscent of the regulatory T-cell 'memory' that is seen for some self-antigens and that helps to prevent autoimmunity⁸.

These fetus-specific memory cells might explain why pre-eclampsia (a condition associated with impaired maternal immune tolerance of the fetus) is predominantly a disease of the first pregnancy unless there is a change in father.

The generation of regulatory T-cell memory towards fetal antigens is part of a broader change in the immunological status quo that is triggered by pregnancy. For example, some autoimmune diseases, such as arthritis, are temporarily ameliorated, and regulatory T cells have been shown to be responsible for this beneficial effect⁹. Unfortunately, the diseases return with a vengeance after delivery and do not seem to benefit from the generation of protective regulatory T-cell memory. However, studies building on Rowe and colleagues' report of regulatory T-cell persistence following pregnancy may in the future help to harness elements of this process for the treatment of autoimmune disease. ■

Alexander G. Betz is at the MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK.
e-mail: betz@mrc-lmb.cam.ac.uk

- Rowe, J., Ertelt, J., Xin, L. & Way, S. S. *Nature* **490**, 102–106 (2012).
- Sasaki, Y. *Mol. Hum. Reprod.* **10**, 347–353 (2004).
- Tafari, A., Alferink, J., Möller, P., Hämmerling, G. J. & Arnold, B. *Science* **270**, 630–633 (1995).
- Aluvihare, V. R., Kallikourdis, M. & Betz, A. G. *Nature Immunol.* **5**, 266–271 (2004).
- Sakaguchi, S., Miyara, M., Costantino, C. M. & Hafler, D. A. *Nature Rev. Immunol.* **10**, 490–500 (2010).
- Andersen, K. G., Nissen, J. K. & Betz, A. G. *Front. Immunol.* **3**, 113 (2012).
- Samstein, R. M., Josefowicz, S. Z., Arvey, A., Treuting, P. M. & Rudensky, A. Y. *Cell* **150**, 29–38 (2012).
- Rosenblum, M. D. *et al. Nature* **480**, 538–542 (2011).
- Munoz-Suano, A., Kallikourdis, M., Sarris, M. & Betz, A. G. *J. Autoimmunity* **38**, J103–J108 (2012).

The oyster genome reveals stress adaptation and complexity of shell formation

Guofan Zhang^{1*}, Xiaodong Fang^{2*}, Ximing Guo^{3*}, Li Li^{1*}, Ruibang Luo^{2,4*}, Fei Xu^{1*}, Pengcheng Yang^{2*}, Linlin Zhang^{1*}, Xiaotong Wang^{1*}, Haigang Qi¹, Zhiqiang Xiong², Huayong Que¹, Yinlong Xie^{2,4}, Peter W. H. Holland⁵, Jordi Paps⁵, Yabing Zhu², Fucun Wu¹, Yuanxin Chen², Jiafeng Wang¹, Chunfang Peng², Jie Meng¹, Lan Yang², Jun Liu¹, Bo Wen², Na Zhang¹, Zhiyong Huang², Qihui Zhu¹, Yue Feng², Andrew Mount⁶, Dennis Hedgecock⁷, Zhe Xu⁸, Yunjie Liu², Tomislav Domazet-Lošo⁹, Yishuai Du¹, Xiaoqing Sun², Shoudu Zhang¹, Binghang Liu^{2,4}, Peizhou Cheng¹, Xuanting Jiang², Juan Li¹, Dingding Fan², Wei Wang¹, Wenjing Fu², Tong Wang¹, Bo Wang², Jibiao Zhang¹, Zhiyu Peng², Yingxiang Li¹, Na Li², Jinpeng Wang¹, Maoshan Chen², Yan He³, Fengji Tan², Xiaorui Song¹, Qiumei Zheng², Ronglian Huang¹, Hailong Yang², Xuedi Du¹, Li Chen², Mei Yang¹, Patrick M. Gaffney¹⁰, Shan Wang³, Longhai Luo², Zhicai She¹, Yao Ming², Wen Huang¹, Shu Zhang², Baoyu Huang¹, Yong Zhang², Tao Qu¹, Peixiang Ni², Guoying Miao¹, Junyi Wang², Qiang Wang¹, Christian E. W. Steinberg¹¹, Haiyan Wang¹, Ning Li², Lumin Qian³, Guojie Zhang², Yingrui Li², Huanming Yang², Xiao Liu¹, Jian Wang², Ye Yin² & Jun Wang^{2,12,13}

The Pacific oyster *Crassostrea gigas* belongs to one of the most species-rich but genomically poorly explored phyla, the Mollusca. Here we report the sequencing and assembly of the oyster genome using short reads and a fosmid-pooling strategy, along with transcriptomes of development and stress response and the proteome of the shell. The oyster genome is highly polymorphic and rich in repetitive sequences, with some transposable elements still actively shaping variation. Transcriptome studies reveal an extensive set of genes responding to environmental stress. The expansion of genes coding for heat shock protein 70 and inhibitors of apoptosis is probably central to the oyster's adaptation to sessile life in the highly stressful intertidal zone. Our analyses also show that shell formation in molluscs is more complex than currently understood and involves extensive participation of cells and their exosomes. The oyster genome sequence fills a void in our understanding of the Lophotrochozoa.

Oceans cover approximately 71% of the Earth's surface and harbour most of the phylum diversity of the animal kingdom. Understanding marine biodiversity and its evolution remains a major challenge. The Pacific oyster *C. gigas* (Thunberg, 1793) is a marine bivalve belonging to the phylum Mollusca, which contains the largest number of described marine animal species¹. Molluscs have vital roles in the functioning of marine, freshwater and terrestrial ecosystems, and have had major effects on humans, primarily as food sources but also as sources of dyes, decorative pearls and shells, vectors of parasites, and biofouling or destructive agents. Many molluscs are important fishery and aquaculture species, as well as models for studying neurobiology, biomineralization, ocean acidification and adaptation to coastal environments under climate change^{2,3}. As the most speciose member of the Lophotrochozoa, phylum Mollusca is central to our understanding of the biology and evolution of this superphylum of protostomes.

As sessile marine animals living in estuarine and intertidal regions, oysters must cope with harsh and dynamically changing environments. Abiotic factors such as temperature and salinity fluctuate wildly, and toxic metals and desiccation also pose serious challenges. Filter-feeding oysters face tremendous exposure to microbial pathogens. Oysters do have a notable physical line of defence against predation and desic-

cation in the formation of thick calcified shells, a key evolutionary innovation making molluscs a successful group. However, acidification of the world's oceans by uptake of anthropogenic carbon dioxide poses a potentially serious threat to this ancient adaptation⁴. Understanding biomineralization and molluscan shell formation is, thus, a major area of interest⁵. *Crassostrea gigas* is also an interesting model for developmental biology owing to its mosaic development with typical molluscan stages, including trochophore and veliger larvae and metamorphosis.

A complete genome sequence of *C. gigas* would enable a more thorough understanding of oyster biology and the evolution of Lophotrochozoa. One of the main challenges, however, is the high levels of polymorphism present in oysters and many marine invertebrates^{6–8}. To overcome this, an oyster derived from four generations of full-sibling mating (coefficient of inbreeding, $F = 0.59$) was used for genome sequencing and assembly (Supplementary Text B1) through fosmid pooling, next-generation sequencing (NGS) and hierarchical assembling. Combining these genomic data with transcriptomes from different organs, different developmental stages and adults challenged with stressors, in addition to mass spectrometric analysis of shell proteins, allowed us to explore characteristics of the oyster genome

¹Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China. ²BGI-Shenzhen, Shenzhen 518083, China. ³Haskin Shellfish Research Laboratory, Institute of Marine and Coastal Sciences, Rutgers University, Port Norris, New Jersey 08349, USA. ⁴HKU-BGI Bioinformatics Algorithms and Core Technology Research Laboratory, Hong Kong. ⁵Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. ⁶Department of Biological Sciences, Clemson University, South Carolina 29634, USA. ⁷Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA. ⁸Atlantic Cape Community College, Mays Landing, New Jersey 08330, USA. ⁹Laboratory of Evolutionary Genetics, Ruđer Bošković Institute, Bijenička cesta 54, P.P. 180, HR-10002, Zagreb, Croatia. ¹⁰School of Marine Science and Policy, University of Delaware, Lewes, Delaware 19958, USA. ¹¹Institute of Biology, Humboldt Universität zu Berlin Arboretum, Späthstraße 80/81, 12437 Berlin, Germany. ¹²Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark. ¹³The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, DK-2200 Copenhagen, Denmark.

*These authors contributed equally to this work.

and key aspects of molluscan biology related to stress response and shell formation.

Sequencing and hierarchical assembly

NGS technology has been successfully applied for *de novo* genome sequencing and assembly using whole-genome shotgun strategies^{9–13}. We initially generated 155-fold Illumina whole-genome shotgun reads (Supplementary Table 1), but could not adequately assemble them owing to high levels of polymorphism and abundant repetitive sequences (Supplementary Text B2 and Supplementary Fig. 1). As possible alternative sequencing strategies—such as the addition of longer Roche 454 reads^{12,13} or traditional bacterial artificial chromosome (BAC)-to-BAC sequencing—are expensive, we opted instead for a more cost-effective fosmid-pooling strategy. In brief, a fosmid library was constructed, and 145,170 clones (~tenfold genome coverage) were evenly and randomly assigned into 1,613 pools, each of which was sequenced to ~60-fold depth and assembled separately (Fig. 1 and Supplementary Table 1). Contigs from each pool were merged into supercontigs, totalling 1,002 megabases (Mb) (Supplementary Text B4.1–3), which was larger than genome-size estimates of 637 Mb from flow cytometry or 545 Mb from *k*-mer (*k*-base fragment) analysis (Supplementary Text B1, 2.3), owing to failure of some allelic variants to merge (Supplementary Figs 3 and 4). Self-to-self whole-genome alignment with LASTZ¹⁴ and sequencing depth information were used to remove redundancy in the assembly (Supplementary Text B4.4). The resulting 446 Mb of the assembly were retained for further scaffolding using paired-end data (Fig. 1). The final assembly comprised 559 Mb, with a contig N50 size (at which 50% of assembly was covered) of 19.4 kilobases (kb) and a scaffold N50 size of 401 kb (Supplementary Text B4.5 and Supplementary Table 3). Over 90% of the assembly was covered by the longest 1,670 (14%) scaffolds.

To assess the completeness of the assembly, 105-fold coverage of short-insert library reads (<2 kb) that participated in assembly (Supplementary Table 1) were aligned against the assembly. Over 99% of these reads were successfully mapped, using a combination

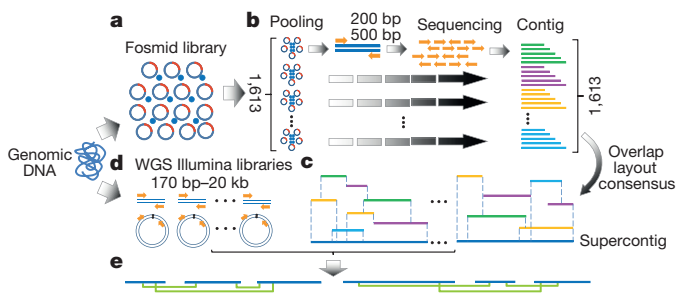


Figure 1 | Fosmid-pooling strategy for oyster genome assembly. Genomic DNA was randomly sheared into fragments. **a**, **b**, A 40-kb-insert fosmid library was constructed (**a**), and 145,170 fosmid clones were randomly selected and assigned into 1,613 pools, each containing 90 clones covering 0.57% of the diploid genome (**b**). For each pool, three Illumina short-insert barcoded libraries (two 200 bp and one 500 bp) were constructed and ~60-fold coverage of 90-bp reads (20-fold per library) were generated, and assembled using SOAPdenovo with optimizing parameters. Assemblies from each pool were further corrected and reassembled if unexpected connections were detected owing to high similarity sequences from different fosmids, and gaps were filled by local assembly. **c**, Fosmid scaffolds were split into contigs at unfilled regions, leaving no undetermined bases in the sequences. Each base was assigned a Phred-like quality score determined by its coverage and alignment mismatches, and these sequences were merged into supercontigs using the overlap layout consensus method. Redundancy was removed using self-to-self alignment and sequencing depth information. **d**, Whole-genome shotgun Illumina libraries (200-bp to 20-kb inserts) from sheared genomic DNA were constructed for mated-pair Illumina sequencing. **e**, The fosmid supercontigs were linked into scaffolds using (1) the whole-genome shotgun sequences; (2) inferred paired-end information extracted from assembled pool scaffolds with a span size ranging from 50 bp to 37.5 kb; and (3) 225,000 fosmid ends sequenced using Sanger technology.

of Burrows–Wheeler Aligners¹⁵ and the more sensitive LASTZ (Supplementary Fig. 5 and Supplementary Table 4). The integrity of the assembly is further demonstrated by the successful mapping of 99% of the BAC sequenced obtained using the Sanger sequencing technique, and 98% of ~68,000 expressed sequence tags from 454 sequencing (Supplementary Text B5, Supplementary Fig. 6 and Supplementary Tables 5 and 6). Fosmid pooling has been used for re-sequencing^{16,17}, and our results show that the combination of fosmid pooling, NGS and hierarchical assembly provides a new, cost-effective alternative for *de novo* sequencing and assembly of complex genomes.

Polymorphism and repetitive sequences

To understand polymorphism in the oyster genome, we analysed allelic variation in the assembled genome (inbred) and one re-sequenced wild oyster (wild) (Supplementary Text C1). The inbred genome contained 3.1 million single-nucleotide polymorphisms and 258,405 short insertion/deletion (indels, 1–40 base pairs (bp)) yielding a sequence polymorphism rate of 0.73%, whereas the wild genome had 3.8 million single-nucleotide polymorphisms and 238,182 indels, or a polymorphism rate of 1.3% (Supplementary Table 7), comparable to previous estimates¹⁸. This 44% reduction in polymorphism in the inbred genome is smaller than the 59.4% predicted from four generations of brother–sister mating, indicating that selection favouring heterozygotes had occurred¹⁹. The polymorphism combining inbred and wild (among four haplotypes) was 2.3%, higher than that in most studied animal genomes^{20,21} but comparable to that in known high-polymorphism species⁷. In inbred and wild, we found 3,094 short indels located in coding regions inferred to cause frameshift variants in 2,665 genes, providing an important source for recessive lethal mutations.

k-mer-based analysis of the oyster genome showed that ~35% of 17-mers had at least two identical copies in the genome, suggesting an abundance of repetitive sequences (Supplementary Fig. 1). Similarly, homology searching and *ab initio* prediction found 202 Mb (36% of the genome) in repetitive sequences (Supplementary Text C2 and Supplementary Table 8). Over 62% of the detected repeats could not be assigned to known categories, reflecting the paucity of genomic information from molluscs²². Large numbers of transposase (359) and reverse-transcriptase (779) gene fragments were detected; over 96% of these had detectable transcripts (Supplementary Fig. 8). Alignment of the wild sequence against the assembly identified 20,605 deletions (>100 bp), over 80% of which overlapped with detected transposable elements, suggesting that transposable elements may have an active role in shaping genome variation. Using MITE-hunter²³, we detected 157,007 copies of miniature inverted-repeat transposable elements (MITEs), accounting for a remarkable 8.82% of the genome (Supplementary Text C2.3 and Supplementary Table 9). Pair-wise comparisons show extremely low sequence divergence in some MITE families (Supplementary Fig. 9), indicating that they may still be active.

Gene annotation and developmental genomics

A total of 28,027 genes were predicted encoding 50 amino acids or more by combining *de novo* prediction and evidence-based searches using reference genomes, oyster expressed sequence tags and transcriptomes from multiple organs and developmental stages (Supplementary Text D1 and E1 and Supplementary Fig. 11), with 96.1% showing expression (reads per kb per million mapped reads (RPKM) > 1 in at least one transcriptome; Supplementary Text D2). Of the inferred proteins, 21,085 matched entries in the SWISS-PROT, InterPro or TrEMBL databases. These genes plus their transcriptome profile from 7 adult organs and at 38 developmental stages provide valuable resources for comparative genomics analysis (Supplementary Text E2 and 3), functional inference and studies on development and organogenesis (Supplementary Text F2 and Supplementary Fig. 15).

One notable finding of developmental interest is that the oyster Hox gene cluster is broken into four sections (Fig. 2) with flanking non-Hox genes (Supplementary Fig. 16). We did not find a clear

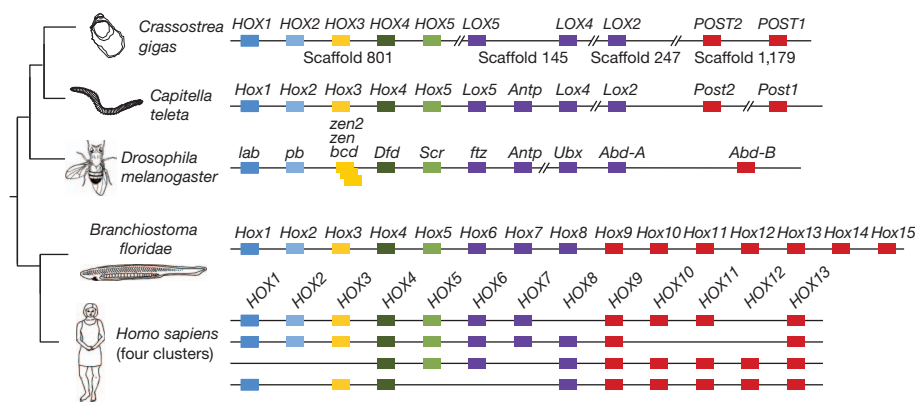


Figure 2 | Clustering of Hox genes in Pacific oyster *Crassostrea gigas*, polychaete annelid *Capitella teleta*, fruitfly *Drosophila melanogaster*, lancelet *Branchiostoma floridae* and *Homo sapiens*. Oblique lines indicate

regions of Hox cluster that are non-contiguous or interrupted. Blue denotes anterior Hox genes, yellow denotes paralogy group 3 Hox genes, green and purple denote central Hox genes and red denotes posterior Hox genes.

Antennapedia gene, which is present in other bivalves such as *Pecten* and *Yoldia*²⁴ (Supplementary Fig. 17). Disruption of the Hox cluster, as also observed in tunicates, nematodes and drosophilids, has been attributed to the loss of temporal co-linearity and modified developmental control²⁵. Supporting this model, we find that Hox genes in the oyster are not activated in an order matching their identity or genomic position, with, for example, *HOX4* and *HOX1* peaking before gastrulation, *LOX5* and *POST2* during the trochophore stage and *HOX5* during the pediveliger stage (Supplementary Fig. 18 and Supplementary Table 15).

Adaptation to environmental stress

Comparison with seven other sequenced genomes identified 8,654 oyster-specific genes (Supplementary Text E3.1) that are probably important in the evolution and adaptation of oysters and other molluscs. With oysters being the only representative, these genes could be shared by other molluscs. Among these genes, gene ontology terms related to 'protein binding', 'apoptosis', 'cytokine activity'

and 'inflammatory response' are highly enriched ($P < 0.0001$; Supplementary Text E2 and Supplementary Table 17), indicating over-representation of some host-defence genes against biotic and abiotic stress. Manual examination shows that several gene families related to defence pathways, including protein folding, oxidation and anti-oxidation, apoptosis and immune responses, are expanded in *C. gigas* (Fig. 3a and Supplementary Table 18). The oyster genome contains 88 heat shock protein 70 (*HSP70*) genes, which have crucial roles in protecting cells against heat and other stresses, compared with ~17 in humans and 39 in sea urchins. Phylogenetic analysis finds clustering of 71 oyster *HSP70* genes to themselves, suggesting that the expansion is specific to the oyster (Supplementary Fig. 19). Also expanded are cytochrome P450 (Supplementary Fig. 20) and multi-copper oxidase gene families, which are important in the biotransformation of endobiotic and xenobiotic chemicals²⁶, and extracellular superoxide dismutases, which are important in defence against oxidative stress. The oyster genome has 48 genes coding for inhibitor of apoptosis proteins (IAPs), compared with 8 in humans and 7 in sea urchins,

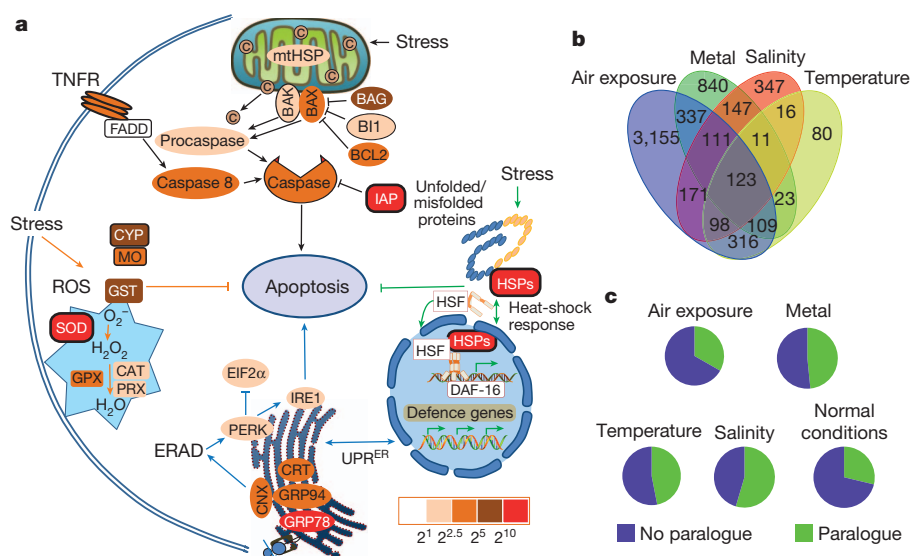


Figure 3 | Expansion, expression and pathway distribution of defence-related genes in *Crassostrea gigas*. **a**, Expansion and expression of key genes in major stress-response pathways in *C. gigas*. Genes include *HSPs* and *HSF* in the heat-shock response; *GRP78*, *CRT*, *CNX*, *GRP94*, *PERK*, *IRE1* and *EIF2 α* in the endoplasmic reticulum unfolded-protein response (*UPR^{ER}*); *IAPs*, *BCL2* like, *BAG*, *BII*, caspases, *FADD* and *TNFR* in apoptotic pathways; *CYP450* and *MO* in oxidation; and *SOD*, *GPX*, *PRX* and *CAT* in anti-oxidation. Boxes with bold black borders indicate gene families (*HSPs*, *IAPs* and *SODs*) expanded in *C. gigas*, and the filled colours correspond to their degree of upregulation in $\text{RPKM}_{\text{treatment}}/\text{RPKM}_{\text{control}}$ by stress, found in 61 transcriptomes from oysters

challenged with 9 types of stressors (Supplementary Text G2 and Supplementary Table 23). **b**, Venn diagram of common and unique genes expressed in response to temperature, salinity, air exposure and heavy-metal stress (zinc, cadmium, copper, lead and mercury), showing overlap of responses. **c**, Number of genes with and without detectable paralogs differentially expressed under stress and normal conditions, showing that genes responding to stress are more likely to have paralogs ($P < 1 \times 10^{-10}$; χ^2 test). Green sections of the pie chart represent 1,442, 809, 358, 550 and 7,938 paralogs for air exposure, metal, temperature, salinity and normal conditions, respectively.

indicating a powerful anti-apoptosis system in oysters. Genes encoding lectin-like proteins, including C-type lectin, fibrinogen-related proteins and C1q domain-containing proteins (C1QDCs), are highly over-represented in the oyster genome ($P < 0.0001$; Supplementary Table 18); these genes have important roles in the innate immune response in invertebrates^{27–29}. Interestingly, many immune-related genes, including genes coding for Gram-negative bacteria-binding proteins, peptidoglycan-recognition proteins, defensin, C-type-lectin-domain-containing proteins and C1QDCs, are highly expressed in the digestive gland (Supplementary Fig. 21), indicating that the digestive system of this filter feeder is an important first-line defence organ against pathogens.

To investigate genome-wide responses to stress, we sequenced 61 transcriptomes from *C. gigas* subjected to nine stressors, including temperature, salinity, air exposure and heavy metals (Supplementary Text G1 and Supplementary Tables 19 and 20). We found that 5,844 genes were differentially expressed under at least one stressor, and genes responding to different stressors showed significant overlap (Fig. 3b and Supplementary Fig. 23a). Air exposure induced a response from the largest number of genes (4,420), indicating that air exposure is a major stressor and that oysters have evolved an extensive gene set in defence. Genes differentially expressed in response to stress are more likely to have paralogues (Fig. 3c), suggesting that expansion and selective retention of duplicated defence-related genes are probably important to oyster adaptation. Under most stressors, genes coding for HSPs, histones, IAPs and protein biogenesis were upregulated, and those for protein degradation downregulated, pointing to concerted responses to maintain cellular homeostasis³⁰ (Supplementary Text G3 and Supplementary Table 21). Genes involved in the unfolded protein response to cellular stress in the endoplasmic reticulum (coding for calreticulin, calnexin, 78- and 94-kDa glucose-regulated proteins) were upregulated, indicating that protein quality control is critical in cellular homeostasis under stress.

Air exposure induced up to 67-fold upregulation of five highly expressed IAPs (Supplementary Fig. 24a). Other inhibitors of apoptosis were also upregulated: *BCL2* up to fourfold and *BAG* up to 12-fold (Supplementary Fig. 24b). These apoptosis inhibitors were also highly upregulated under heat and low salinity stress. These findings, along with the expansion of IAPs, suggest that a powerful anti-apoptosis system exists and may be critical for the amazing endurance of oysters to air exposure and other stresses. The existence of an intrinsic apoptosis pathway in invertebrates has been controversial, and parts of the pathways have only recently been demonstrated for two lophotrochozoans^{31,32}. The finding of key genes belonging to both intrinsic (*BAX*, *BAK*, *BAG*, *BCL2*, *BII* and procaspase) and extrinsic (*TNFR* and caspase 8) apoptosis pathways indicates that oysters have advanced apoptosis systems. Powerful inhibition of apoptosis as shown by genomic and transcriptomic analyses may be central to the ability of oysters to tolerate prolonged air exposure and other stresses.

Heat stress induced a ~2,000-fold increase in expression of five highly inducible *HSP70* genes or a 13.9-fold increase in average expression of all *HSP70* genes, amounting to 4.2% of all transcripts (Supplementary Figs 24c and 25). The genomic expansion and massive upregulation of HSP genes help to explain why *C. gigas* can tolerate temperatures as high as 49 °C when exposed to summer sun at low tide³³. HSP genes were also upregulated under other stressors and may be central to the oyster defence against all stresses (Supplementary Fig. 25). HSP genes may also inhibit apoptosis by binding to effector caspases³⁴.

Genes involved in signal transduction, including genes coding for G-protein-coupled receptors and Ras GTPase, were also activated by stressors (Supplementary Fig. 24f) and over-represented in the oyster genome (Supplementary Table 11). These regulators may have a role in orchestrating stress responses, which seem to be well coordinated (Fig. 3a and Supplementary Fig. 25). The expansion of key defence genes and the strong, complex transcriptomic response to

stress highlight the sophisticated genomic adaptations of the oyster to sessile life in a highly stressful environment.

Shell formation

Calcified shells provide critical protection against predation and desiccation in sessile marine animals such as oysters. Molluscan shells consist of calcium carbonate (CaCO_3) crystals of either aragonite or calcite embedded in an elaborate organic matrix. Two models have been advanced for molluscan shell formation. The matrix model posits that mineralization occurs in a mantle-secreted matrix of chitin, silk fibroin and acidic proteins^{35,36}. Chitin and silk proteins are proposed to provide matrix structure, whereas acidic proteins control the nucleation and growth of CaCO_3 crystals. The cellular model suggests that biomineralization is cell-mediated; that is, crystals are formed in haemocytes and then deposited at the mineralization front³⁷.

We searched the oyster genome for genes implicated in shell formation in previous studies and examined their expression in different tissues and at different stages (Supplementary Text H1, 2). We also sequenced peptides from shells, mapped them to the genome and identified 259 shell proteins (Supplementary Text H3 and Supplementary Table 24). Although our search found evidence for the involvement of chitin, we did not find any silk-like proteins encoded in the oyster genome (Supplementary Text H2) but found, instead, many diverse proteins that may have roles in matrix construction and modification. Notably, a gene coding for a fibronectin-like protein was highly expressed at the early developmental stage, when larval shells are formed, in unison with chitin synthase (Fig. 4a) and was mostly expressed in the adult mantle (40× other organ average; Fig. 4b); the fibronectin-like protein was among the most abundant proteins found in oyster shells. Genes coding for laminin and some collagen proteins were also highly expressed in the mantle (Supplementary Fig. 27a) and found in shells. These are typical extracellular matrix (ECM) proteins, and their presence in shells suggests that the shell matrix has similarities to the ECM of animal connective tissues and basal lamina. Unlike silk fibroins that can self assemble³⁸, the formation of fibronectin fibrils in the ECM is cell mediated³⁹. Oyster fibronectin-like proteins have five type-III domains for integrin binding and cell adhesion. Genes coding for integrins were highly expressed in haemocytes (4× other organ average, Supplementary Fig. 27b). Thus, haemocytes may organize fibronectin-like fibril formation in the shell matrix as they do in ECM.

The involvement of cells in shell formation is further supported by the functional diversity of proteins detected in shells. Many house-keeping proteins, such as elongation factor 1 α and ribosomal proteins, were found in the shell; indeed, most oyster shell proteins are not structural proteins but are distributed in diverse metabolic pathways (Fig. 4c and Supplementary Table 25). This functional diversity of shell proteins mirrors that of cells, which is unexpected under the matrix model. Furthermore, 84% of the 259 shell proteins identified are not classical secreted proteins (Supplementary Text H3.4 and Supplementary Table 24); they may be part of cells or deposited by exosomes⁴⁰. Supporting the presence of exosomes, 61 of the 259 shell proteins matched proteins in the exosome database⁴¹. Cells and exosome-like vesicles containing calcite crystals have been observed at the mineralization front^{37,42}, although their significance in shell formation is debated. This study provides molecular evidence for their presence inside shells and their probable participation in shell formation.

Many shell proteins are enzymes that may be involved in matrix construction or modification. A homologue of penicillin-binding protein is exclusively expressed in mantle (72× other organ average) and highly abundant in shells (Supplementary Fig. 27d). Penicillin-binding protein is a transpeptidase that crosslinks glycopeptides in bacterial cell walls⁴³ and may have similar functions in the shell matrix. Another notable enzyme found is tyrosinase. The oyster

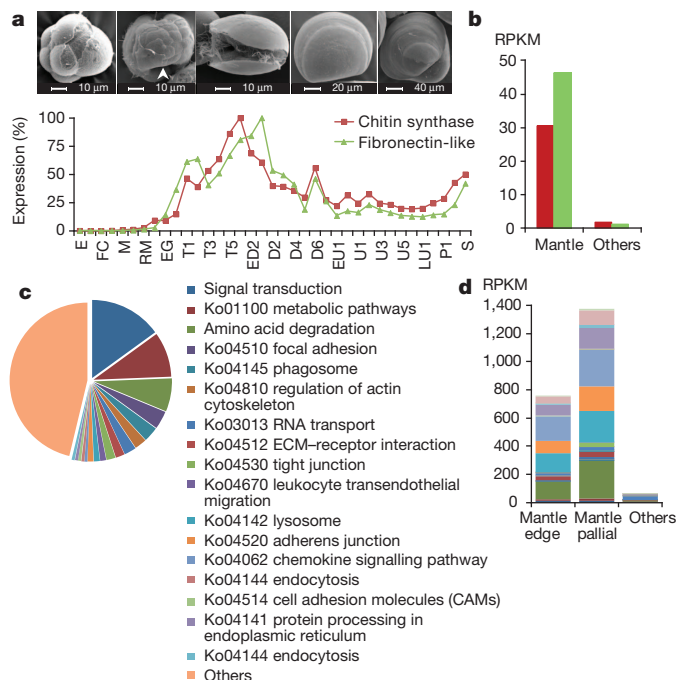


Figure 4 | Genes related to shell formation identified from mass spectroscopy analysis of shell proteins and transcriptome data. **a**, Relative expression (y axis) of genes coding for chitin synthase (gene CGI_10009438) and fibronectin-like (CGI_10016964) in early development corresponds to the formation of shell gland and first larval shells, as seen in scanning electron microscope photos. White arrow denotes the invagination that forms the shell gland. Developmental stages (x axis) and their timeline are defined in Supplementary Table 12. **b**, In adults, chitin synthase and fibronectin-like proteins (same colour as in **a**) are almost exclusively expressed in the mantle compared with other organs. Fibronectin-like is also one of the most abundant proteins found in the shell. **c**, Distribution of shell proteins in diverse Kyoto encyclopedia of genes and genomes (KEGG) pathways indicative of general cellular functions. **d**, Expression of 26 tyrosinase genes in the mantle edge, mantle pallial and other organs. Tyrosinases are abundant in shells and their higher expression in the non-pigmented mantle pallial indicate that their functions are not limited to melanogenesis but are related to shell formation.

genome has an expanded set of 26 genes coding for tyrosinase, compared with one in *Caenorhabditis elegans* and two in humans; most genes coding for tyrosinase are mantle specific (Fig. 4d) and highly enriched among shell proteins ($P = 8 \times 10^{-6}$). Although tyrosinase is a key enzyme in melanogenesis^{44,45}, it is most highly expressed in the non-pigmented pallial mantle (Fig. 4d), indicating that it has other functions in the oyster. The mantle secretes tyrosine-rich proteins⁴⁶, and oxidation of tyrosine may be essential for shell matrix maturation. Several proteinases and proteinase inhibitors are highly mantle specific and abundant in shells, and may be involved in matrix formation, modification and protection (Supplementary Table 24). Together, these results indicate that oyster shell matrix is not formed simply by self-assembling silk-like proteins but by diverse proteins through complex assembly and modification processes that may involve haemocytes and exosomes.

Concluding remarks

We sequenced and assembled the genome of the Pacific oyster using an inbred individual, short-read NGS and a new fosmid-pooling and hierarchical assembly strategy. The draft assembly provided insight into a molluscan genome characterized by high polymorphism, abundant repetitive sequences and active transposable elements. Genomic, transcriptomic and proteomic analyses show unique adaptations of oysters to sessile life in a highly stressful intertidal environment and the complexity of shell formation. The oyster

genome sequence and comprehensive transcriptome data provide valuable resources for studying molluscan biology and lophotrochozoan evolution, and for genetic improvement of oysters and other important aquaculture species.

METHODS SUMMARY

The sequenced Pacific oyster is an inbred female produced by four generations of brother–sister mating. Genome sequences were produced with Illumina platform using fosmid pooling and assembled with a new hierarchical assembly strategy. Fosmid ends were sequenced by Sanger. Gene models were obtained by integrating results of *de novo* gene prediction and alignment-based methods based on homology and transcriptomic evidence. Transcriptomes were sequenced with Illumina platform. The proteome of the shell was obtained by mass spectrometry. All methods are described in detail in the Supplementary Information.

Received 30 January; accepted 11 July 2012.

Published online 19 September; corrected online 3 October 2012.

- Ponder, W. F. & Lindberg, D. R. in *Phylogeny and Evolution of the Mollusca* (eds Ponder, W. & Lindberg, D. R.) Ch. 1, 1–18 (Univ. of California Press, 2008).
- Walters, E. T. & Moroz, L. L. Molluscan memory of injury: evolutionary insights into chronic pain and neurological disorders. *Brain Behav. Evol.* **74**, 206–218 (2009).
- Talmage, S. C. & Gobler, C. J. Effects of past, present, and future ocean carbon dioxide concentrations on the growth and survival of larval shellfish. *Proc. Natl Acad. Sci. USA* **107**, 17246–17251 (2010).
- Caldeira, K. & Wickett, M. E. Anthropogenic carbon and ocean pH. *Nature* **425**, 365–365 (2003).
- Marin, F., Luquet, G., Marie, B. & Medakovic, D. Molluscan shell proteins: primary structure, origin, and evolution. *Curr. Top. Dev. Biol.* **80**, 209–276 (2008).
- Hedgecock, D. *et al.* The case for sequencing the Pacific oyster genome. *J. Shellfish Res.* **24**, 429–441 (2005).
- Sodergren, E. *et al.* The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **314**, 941–952 (2006).
- Small, K. S., Brudno, M., Hill, M. M. & Sidow, A. Extreme genomic variation in a natural population. *Proc. Natl Acad. Sci. USA* **104**, 5698–5703 (2007).
- Li, R. *et al.* The sequence and *de novo* assembly of the giant panda genome. *Nature* **463**, 311–317 (2010).
- Xu, X. *et al.* Genome sequence and analysis of the tuber crop potato. *Nature* **475**, 189–195 (2011).
- Bonasio, R. *et al.* Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329**, 1068–1071 (2010).
- Dalloul, R. A. *et al.* Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol.* **8**, e1000475 (2010).
- Star, B. *et al.* The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477**, 207–210 (2011).
- Harris, R. S. *Improved Pairwise Alignment of Genomic DNA*. PhD thesis, Pennsylvania State Univ. (2007).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Kitzman, J. O. *et al.* Haplotype-resolved genome sequencing of a Gujarati Indian individual. *Nature Biotechnol.* **29**, 59–63 (2010).
- Suk, E. K. *et al.* A comprehensively molecular haplotype-resolved genome of a European individual. *Genome Res.* **21**, 1672–1685 (2011).
- Sauvage, C., Bierre, N., Lapègue, S. & Boudry, P. Single nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. *Gene* **406**, 13–22 (2007).
- McGoldrick, D. J. & Hedgecock, D. Fixation, segregation and linkage of allozyme loci in inbred families of the Pacific oyster *Crassostrea gigas* (Thunberg): implications for the causes of inbreeding depression. *Genetics* **146**, 321–334 (1997).
- Hillier, L. W. *et al.* Whole-genome sequencing and variant discovery in *C. elegans*. *Nature Methods* **5**, 183–188 (2008).
- Li, R. *et al.* *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010).
- Gaffney, P. M., Pierce, J. C., Mackinley, A. G., Titchen, D. A. & Glenn, W. K. *Pearl*, a novel family of putative transposable elements in bivalve mollusks. *J. Mol. Evol.* **56**, 308–316 (2003).
- Han, Y. & Wessler, S. R. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**, e199 (2010).
- Barucca, M., Olmo, E. & Canapa, A. Hox and paraHox genes in bivalve molluscs. *Gene* **317**, 97–102 (2003).
- Ferrier, D. E. K. & Holland, P. W. H. *Ciona intestinalis* ParaHox genes: evolution of Hox/ParaHox cluster integrity, developmental mode, and temporal colinearity. *Mol. Phylogenet. Evol.* **24**, 412–417 (2002).
- Goldstone, J. V. *et al.* The chemical defenseome: environmental sensing and response genes in the *Strongylocentrotus purpuratus* genome. *Dev. Biol.* **300**, 366–384 (2006).
- Carland, T. M. & Gerwick, L. The C1q domain containing proteins: where do they come from and what do they do? *Dev. Comp. Immunol.* **34**, 785–790 (2010).
- Hanington, P. C. & Zhang, S. M. The primary role of fibrinogen-related proteins in invertebrates is defense, not coagulation. *J. Innate Immun.* **3**, 17–27 (2011).

29. Zhang, S. M., Adema, C. M., Kepler, T. B. & Loker, E. S. Diversification of Ig superfamily genes in an invertebrate. *Science* **305**, 251–254 (2004).
30. Kourtis, N. & Tavernarakis, N. Cellular stress response pathways and ageing: intricate molecular relationships. *EMBO J.* **30**, 2520–2531 (2011).
31. Lee, E. F. *et al.* Discovery and molecular characterization of a Bcl-2-regulated cell death pathway in schistosomes. *Proc. Natl Acad. Sci. USA* **108**, 6999–7003 (2011).
32. Bender, C. E. *et al.* Mitochondrial pathway of apoptosis is ancestral in metazoans. *Proc. Natl Acad. Sci. USA* **109**, 4904–4909 (2012).
33. Galtsoff, P. S. The American oyster *Crassostrea virginica* Gmelin. *Fishery Bull.*, **64**, 1–480 (United States Govt Printing Office, 1964).
34. Mosser, D. D., Caron, A. W., Bourget, L., Denis-Larose, C. & Massie, B. Role of the human heat shock protein hsp70 in protection against stress-induced apoptosis. *Mol. Cell. Biol.* **17**, 5317–5327 (1997).
35. Weiner, S., Traub, W. & Parker, S. Macromolecules in mollusc shells and their functions in biomineralization. *Phil. Trans. R. Soc. Lond. B* **304**, 425–434 (1984).
36. Furuhashi, T., Schwarzing, C., Miksik, I., Smrz, M. & Beran, A. Molluscan shell evolution with review of shell calcification hypothesis. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **154**, 351–371 (2009).
37. Mount, A. S., Wheeler, A. P., Paradkar, R. P. & Snider, D. Hemocyte-mediated shell mineralization in the eastern oyster. *Science* **304**, 297–300 (2004).
38. Stark, M. *et al.* Macroscopic fibers self-assembled from recombinant miniature spider silk proteins. *Biomacromolecules* **8**, 1695–1701 (2007).
39. Lemmon, C. A., Chen, C. S. & Romer, L. H. Cell traction forces direct fibronectin matrix assembly. *Biophys. J.* **96**, 729–738 (2009).
40. Keller, S., Sanderson, M. P., Stoeck, A. & Altevogt, P. Exosomes: from biogenesis and secretion to biological function. *Immunol. Lett.* **107**, 102–108 (2006).
41. Mathivanan, S., Fahner, C. J., Reid, G. E. & Simpson, R. J. ExoCarta 2012: database of exosomal proteins, RNA and lipids. *Nucleic Acids Res.* **40**, D1241–D1244 (2012).
42. Mount, A. S., Gohad, N. V., Hansen, D. C., Mueller, K. & Johnstone, M. B. Deposition of nanocrystalline calcite on surfaces by a tissue and cellular biomineralization. US patent 2010/0150982 A1 (2010).
43. Sauvage, E., Kerff, F., Terrak, M., Ayala, J. A. & Charlier, P. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol. Rev.* **32**, 234–258 (2008).
44. Nagai, K., Yano, M., Morimoto, K. & Miyamoto, H. Tyrosinase localization in mollusc shells. *Comp. Biochem. Physiol. B* **146**, 207–214 (2007).
45. Chang, T. S. An updated review of tyrosinase inhibitors. *Int. J. Mol. Sci.* **10**, 2440–2475 (2009).
46. Waite, J. H. in *The Mollusca* Vol. I (eds Hochachka, P. & Wilbur, K. M.) Ch. 11, 467–504 (Academic, 1983).
47. Zhang, G. *et al.* Genomic data from the Pacific oyster (*Crassostrea gigas*). GigaScience. <http://dx.doi.org/10.5524/100030> (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge H. Wu, F. Zhang, Q. Tang, Z. Zhu, X. Xu, H. Lin, J. Lei, Z. Xiang, N. Li, J. Xiang and J. Jia for their support of the oyster genome project. We thank F. Han, X. Liu, R. Wu, L. Wang, Y. Wu, L. Yan, H. Niu, H. Li, Y. Wang, J. Liang, Z. Jia, J. Davis and Taylor Shellfish Farms for assistance with DNA, RNA and protein extraction, data analysis and oyster culture, and Y. Lu, C. Lin, H. Peng, Y. Ren, X. Xu, R. Chen and D. Zhang for library construction and sequencing. We thank L. Song, B. Z. Liu, Q. Li, Z. Yu, C. Ke, J. Yu, B. Liu, X. Sun, R. W. Chapman, Y. Han, S. R. Wessler, D. Arendt, E. H. Davidson,

J. S. Evans, B. Brown, P. Boudry and B. Lieb for discussions. We thank other faculty and staff at the Institute of Oceanology, Chinese Academy of Science, BGI-Shenzhen and Rutgers who contributed to the oyster genome project. We acknowledge grant support from the National High-Technology Research and Development Program of China (863 program; 2010AA10A110), National Basic Research Program of China (973 Program; 2010CB126401 and 2010CB126402), 863 program (2012AA10A405), Basic Research Program Supported by Shenzhen City (JC2010526019), Shenzhen Key Laboratory of Transomics Biotechnologies (CXB201108250096A), National Natural Science Foundation of China (40730845), Mollusc Research and Development Center, CARS, Shenzhen Key Laboratory of Gene Bank for National Life Science, Taishan Scholar and Scholar Climbing Programs of Shandong. X.G. acknowledges funding from the US Department of Agriculture (2009-35205-05052 and NJ32108) and the Chinese Academy of Science Marine Functional Genomics Oversea Team and Taishan Scholar Fund; P.W.H.H. acknowledges funding from the European Research Council (EU FP7 ERC grant [268513]11); and J.P. acknowledges funding from Beatrice de Pinós of the Generalitat de Catalunya (2009 BP-DGR). We are grateful to Dalian Zhangzidao Fishery Group Co. Ltd for providing support.

Author Contributions G.Z. and X.G. conceived the study and designed scientific objectives. G.Z., Jun W. and X.G. led the project and manuscript preparation. Jun W., X.F. and Y.Y. developed the sequencing strategy. L. Li and X.F. managed the project. R.L. (leader), Yr.L., Z.H., Y.L., Xq.S., B.L., X.J., W.F., Qm.Z., H.Y., L. Luo, B. Wang, Y.M. and P.N. conducted assembly and evaluation; X.F. (leader), P.Y., Zq.X., Y.X., Yb.Z., Y.C., C.P., Y.F., D.F., L.Y., Z.P., Na L., X.W., M.C., L.C., S.Z., Jy.W., Ning L., Gj.Z. and Yr.L. performed genome annotation and data analysis; L. Li, F.X., Hy.Q., F.W., Sd.Z., Jp.W., X.D., J.Z., Q.W. and L.Q. cultured oysters and provided materials; Hg.Q. (leader), L. Li, Jf.W., Z.S. and H.W. performed polymorphism analysis and validation; F.X. (leader), P.W.H.H., J.P., T.D.L., P.Y., J. Liu, X.W., L. Li, N.Z., J. Li, W.W., Yx.L., M.Y. and W.H. conducted developmental biology studies and data analysis; L.Z. (leader), X.G., J.M., Qh.Z., Y.D., C.E.W.S., P.C., B.H., T.Q. and G.M. conducted stress studies and data analysis; X.W. (leader), X.G., T.W., Z. Xu, Y.H., A.M., Xr.S., R.H., B. Wen, F.T. and Y.Z. conducted shell-formation studies and data analysis. Hm.Y. and Jian W. supervised sequencing, assembly and bioinformatics analysis. X.G., S.W. and F.X. performed flow-cytometry analysis. D.H. and P.M.G. provided inbred oysters, BAC sequences and advice. L.Q. and X.L. participated in discussions and provided suggestions. X.G., X.F., L. Li, R.L., F.X., P.Y., L.Z., X.W., Hg.Q. and P.W.H.H. did most of the writing with contributions from all authors.

Author Information The oyster genome project has been deposited at DDBJ/EMBL/GenBank under the accession number AFT100000000. All short-read data have been deposited into the Sequence Read Archive (SRA) (<http://www.ncbi.nlm.nih.gov/sra>) under the accession number SRA040229. Short-read data of re-sequencing have been deposited in the SRA under the accession number SRA043580. Raw sequencing data of transcriptomes have been deposited in the Gene Expression Omnibus under the accession number GSE31012. Genomic data are also available at the Comprehensive Library for Modern Biotechnology (CLIMB) repository: doi:10.5524/100030 (ref. 47). Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and the online version of the paper is freely available to all readers. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Jun W. (wangj@genomics.org.cn), X.G. (xguo@hsrl.rutgers.edu), Y.Y. (yinye@genomics.org.cn) or G.Z. (gzhang@qdio.ac.cn).

A metagenome-wide association study of gut microbiota in type 2 diabetes

Junjie Qin^{1*}, Yingrui Li^{1*}, Zhiming Cai^{2*}, Shenghui Li^{1*}, Jianfeng Zhu^{1*}, Fan Zhang^{3*}, Suisha Liang¹, Wenwei Zhang¹, Yuanlin Guan¹, Dongqian Shen¹, Yangqing Peng¹, Dongya Zhang¹, Zhuye Jie¹, Wenxian Wu¹, Youwen Qin¹, Wenbin Xue¹, Junhua Li¹, Lingchuan Han³, Donghui Lu³, Peixian Wu³, Yali Dai³, Xiaojuan Sun², Zesong Li², Aifa Tang², Shilong Zhong⁴, Xiaoping Li¹, Weineng Chen¹, Ran Xu¹, Mingbang Wang¹, Qiang Feng¹, Meihua Gong¹, Jing Yu¹, Yanyan Zhang¹, Ming Zhang¹, Torben Hansen⁵, Gaston Sanchez⁶, Jeroen Raes^{7,8}, Gwen Falony^{7,8}, Shujiro Okuda^{7,8}, Mathieu Almeida⁹, Emmanuelle LeChatelier⁹, Pierre Renault⁹, Nicolas Pons⁹, Jean-Michel Batto⁹, Zhaoxi Zhang¹, Hua Chen¹, Ruifu Yang^{1,10}, Weimou Zheng¹, Songgang Li¹, Huanming Yang¹, Jian Wang¹, S. Dusko Ehrlich⁹, Rasmus Nielsen⁶, Oluf Pedersen^{5,11,12}, Karsten Kristiansen^{1,13} & Jun Wang^{1,5,13}

Assessment and characterization of gut microbiota has become a major research area in human disease, including type 2 diabetes, the most prevalent endocrine disease worldwide. To carry out analysis on gut microbial content in patients with type 2 diabetes, we developed a protocol for a metagenome-wide association study (MGWAS) and undertook a two-stage MGWAS based on deep shotgun sequencing of the gut microbial DNA from 345 Chinese individuals. We identified and validated approximately 60,000 type-2-diabetes-associated markers and established the concept of a metagenomic linkage group, enabling taxonomic species-level analyses. MGWAS analysis showed that patients with type 2 diabetes were characterized by a moderate degree of gut microbial dysbiosis, a decrease in the abundance of some universal butyrate-producing bacteria and an increase in various opportunistic pathogens, as well as an enrichment of other microbial functions conferring sulphate reduction and oxidative stress resistance. An analysis of 23 additional individuals demonstrated that these gut microbial markers might be useful for classifying type 2 diabetes.

Type 2 diabetes (T2D), which is a complex disorder influenced by both genetic and environmental components, has become a major public health issue throughout the world^{1,2}. Currently, research to parse the underlying genetic contributors to T2D is mainly through the use of genome-wide association studies (GWAS) focusing on identifying genetic components in the organism's genome^{3,4}. Recently, research has indicated that the risk of developing T2D may also involve factors from the 'other genome', that is, the 'intestinal microbiome' (also termed the gut metagenome)⁵.

Previous metagenomic research on the gut metagenome, primarily using 16S ribosomal RNA⁶ and whole-genome shotgun (WGS) sequencing⁷, has provided an overall picture of commensal microbial communities and their functional repertoire. For example, a catalogue of 3.3 million human gut microbial genes were established in 2010 (ref. 8) and, of note, a more extensive catalogue of gut microorganisms and their genes were published later^{9,10}. Recent research on the gut metagenome has changed our understanding of human disease and its potential medical impact as many studies have reported. From the perspective of both taxonomic and functional composition, the gut microbiota might be linked to and contribute to many complex diseases¹¹. For example, several studies have indicated that obesity is associated with an increase in the phylum Firmicutes and a relatively lower abundance of the phylum Bacteroidetes^{7,12–16}. Crohn's disease research has revealed that patients had a significant reduction in the overall diversity of the gut microbiota¹⁷ and had changes in

microbial composition¹⁸, and a T2D study showed that the proportion of the phylum Firmicutes and the class Clostridia in the gut of patients was significantly reduced¹⁹. However, more work is required to gain detailed information about gut microbial compositional changes and their associated impact with these types of diseases, and additional tools are required to find ways to determine associated changes easily and rapidly.

To reach these initial goals, we devised and carried out a two-stage case-control metagenome-wide association study (MGWAS) based on deep next-generation shotgun sequencing of DNA extracted from the stool samples from a total of 345 Chinese T2D patients and non-diabetic controls. From this we pinpointed specific genetic and functional components of the gut metagenome associated with T2D (Supplementary Fig. 1). Our data provide insight into the characteristics of the gut metagenome related to T2D risk, a paradigm for future studies of the pathophysiological role of the gut metagenome in other relevant disorders, and the potential usefulness for a gut-microbiota-based approach for assessment of individuals at risk of such disorders.

Construction of a gut metagenome reference

To identify metagenomic markers associated with T2D, we first developed a comprehensive metagenome reference gene set that included genetic information from Chinese individuals and T2D-specific gut microbiota, as the currently available metagenomic reference (the MetaHIT gene catalogue) did not include such data. We

¹BGI-Shenzhen, Shenzhen 518083, China. ²Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University, Shenzhen 518035, China. ³Peking University Shenzhen Hospital, Shenzhen 518036, China. ⁴Medical Research Center of Guangdong General Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, China. ⁵The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health Sciences, University of Copenhagen, DK-2100 Copenhagen, Denmark. ⁶Department of Integrative Biology and Department of Statistics, University of California Berkeley, Berkeley, CA 94820, USA. ⁷Department of Structural Biology, VIB, 1050 Brussels, Belgium. ⁸Department of Applied Biological Sciences (DBIT), Vrije Universiteit Brussel, 1050 Brussels, Belgium. ⁹Institut National de la Recherche Agronomique, 78350 Jouy en Josas, France. ¹⁰State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, China. ¹¹Institute of Biomedical Sciences, University of Copenhagen & Faculty of Health Science, University of Aarhus, DK-8000 Aarhus, Denmark. ¹²Hagedorn Research Institute, DK-2820 Gentofte, Denmark. ¹³Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark.

*These authors contributed equally to this work.

carried out WGS sequencing on individual faecal DNA samples from 145 Chinese individuals (71 cases and 74 controls, Supplementary Table 1) and obtained an average of 2.61 gigabases (Gb) (15.8 million) paired-end reads for each, totalling 378.4 Gb of high-quality data that was free of human DNA and adaptor contaminants (Supplementary Table 2). We then performed *de novo* assembly and metagenomic gene prediction for all 145 samples. We integrated these data with the MetaHIT gene catalogue, which contained 3.3 million genes that were predicted from the gut metagenomes of individuals of European descent, and obtained an updated gene catalogue with 4,267,985 predicted genes. A total of 1,090,889 of these genes were uniquely assembled from our Chinese samples, which contributed 10.8% additional coverage of sequencing reads when comparing our data against that from the MetaHIT gene catalogue alone (Supplementary Fig. 2).

Having a more complete gene reference, we carried out taxonomic assignment and functional annotation for the updated gene catalogue using 2,890 reference genomes (IMG v3.4; Supplementary Table 3), KEGG (Release 59.0) and eggNOG databases (v3). Here, 21.3% of the genes in the updated catalogue could be robustly assigned to a genus, which covered 26.4%–90.6% (61.2% on average) of the sequencing reads in the 145 samples (Supplementary Methods); the remaining genes were likely to be from currently undefined microbial species. For assessment at a functional level, we identified 6,313 KEGG orthologues and 38,641 eggNOG orthologue groups in the updated gene catalogue, which covered 47.1% and 60.9%, respectively, of the genes in the catalogue. In addition, 14.0% of genes that were not mapped to eggNOG orthologue groups could be clustered into 7,042 novel gene families; however, these do not yet have any functional annotation information, but were still included (as in-house eggNOG orthologue groups) in our analyses. For each metagenomic sample, on average, 48.7% and 68.8% sequencing reads were covered, respectively, by these KEGG orthologues- and eggNOG orthologue groups-annotated genes.

Marker identification using a two-stage MGWAS

To define T2D-associated metagenomic markers, we devised and carried out a two-stage MGWAS strategy. Using a sequence-based profiling method, we quantified the gut microbiota in the 145 samples for use in stage I. On average, with the requirement that there should be $\geq 90\%$ identity, we could uniquely map $77.4 \pm 0.6\%$ (mean \pm s.e.m.; $n = 145$) paired-end reads to the updated gene catalogue (Supplementary Fig. 2 and Supplementary Table 2). To normalize the sequencing coverage, we used relative abundance instead of the raw read count to quantify the gut microbial genes (Supplementary Methods). With nearly 16 million sequencing reads on average per sample, our sequence-based profiling method could reliably detect very low-abundance genes. For example, given a gene with a real relative abundance of 1×10^{-6} , the detected value ranged from 0.7×10^{-6} to 1.5×10^{-6} based on a theoretical estimation (Supplementary Fig. 3). To facilitate the subsequent statistical analyses at both genetic and functional levels, we further defined and prepared three types of profiles using the quantified gene results: (1) a gene profile; (2) a KEGG orthologues profile; and (3) an eggNOG orthologue groups profile (Supplementary Methods).

We investigated the subpopulations of the 145 samples in these different profiles. Applying the same identification method as used in the MetaHIT study²⁰, we identified three enterotypes in our Chinese samples (Supplementary Figs 4 and 5). A principal component analysis (PCA) showed that these three enterotypes were primarily made up of several highly abundant genera, including *Bacteroides*, *Prevotella*, *Bifidobacterium* and *Ruminococcus* (Fig. 1a). However, we found no significant relationship between enterotype and T2D disease status ($P = 0.29$, Fisher's exact test). We examined the top five principal components (P value in Tracy–Widom test < 0.05 and contribution $> 3\%$): the first and second principal components were significantly correlated with enterotype ($P < 0.001$, Kruskal–Wallis test), and the fifth principal component was significantly correlated with

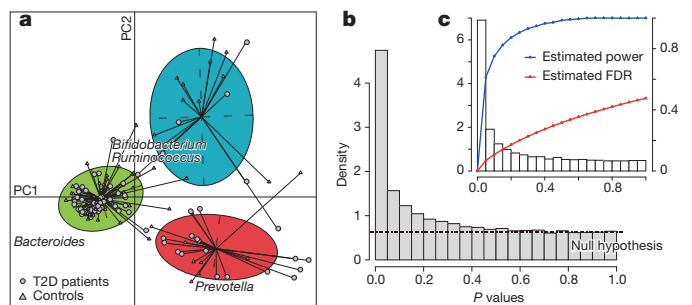


Figure 1 | Identification of T2D-associated markers from gut metagenome.

a, The T2D patients ($n = 71$) and controls ($n = 74$) from stage I were plotted on the first two principal components of the genus profile. Lines connect individuals determined to have the same enterotype (using the PAM clustering method of refs 20,36), and coloured circles cover the individuals near the centre of gravity for each cluster ($< 1.5\sigma$). The top four genera as the main contributors to these clusters were determined and plotted by their loadings in these two components. **b**, Density histogram showing the P -value distribution of all genes tested in stage I. The horizon line represents the distribution of P values under the null hypothesis. **c**, Density histogram showing the P -value distribution of genes in stage II, which were identified from stage I. The blue and red curves denote the estimated statistical power and false discovery rate (FDR), respectively, for a particular P value.

T2D ($P < 0.001$, Wilcoxon rank-sum test; Supplementary Fig. 5d), indicating that T2D, in addition to enterotype, was a determining factor in explaining the gut microbial differences in our samples. The third and fourth principal components, however, did not correlate with any known factors.

We then corrected for population stratification, which might be related to the non-T2D-related factors. For this we analysed our data using a modified EIGENSTRAT method²¹; however, unlike what is done in a GWAS subpopulation correction, we applied this analysis to microbial abundance rather than to genotype. For gene profile, after adjustment, we found that the effects that correlated with non-T2D-related factors disappeared (Supplementary Table 4). A Wilcoxon rank-sum test was done on the adjusted gene profile to identify differential metagenomic gene content between the T2D patients and controls. The outcome of our analyses showed a substantial enrichment of a set of microbial genes that had very small P values, as compared with the expected distribution under the null hypothesis (Fig. 1b), indicating that these genes were true T2D-associated gut microbial genes.

To validate the significant associations identified in stage I, we carried out the stage II analysis using an additional 200 Chinese individuals (one of these samples had a very low within-sample diversity, which was probably owing to the presence of a high fraction of *Escherichia* and *Klebsiella*, and was therefore excluded in later analyses; Supplementary Tables 1 and 2). We also used WGS sequencing in stage II and generated a total of 830.8 Gb sequence data with 23.6 million paired-end reads on average per sample. We then assessed the 278,167 stage I genes that had P values < 0.05 and found that the majority of these genes still correlated with T2D in these stage II study samples (Supplementary Fig. 6). We next controlled for the false discovery rate (FDR) in the stage II analysis, and defined a total of 52,484 T2D-associated gene markers from these genes corresponding to a FDR of 2.5% (stage II P value < 0.01 ; Fig. 1c, Supplementary Fig. 7 and Supplementary Table 5).

We applied the same two-stage analysis using the KEGG orthologues and eggNOG orthologue groups profiles and identified a total of 1,345 KEGG orthologues markers (stage II $P < 0.05$ and 4.5% FDR) and 5,612 eggNOG orthologue groups markers (stage II $P < 0.05$ and 6.6% FDR) that were associated with T2D (Supplementary Tables 6 and 7).

Development of a metagenomic linkage group

To reduce and structurally organize the abundant metagenomic data and to enable us to make a taxonomic description, we devised the

generalized concept of metagenomic linkage group (MLG) in lieu of a species concept for a metagenome. Here a MLG is defined as a group of genetic material in a metagenome that is probably physically linked as a unit rather than being independently distributed; this allowed us to avoid the need to completely determine the specific microbial species present in the metagenome, which is important given there are a large number of unknown organisms and that there is frequent lateral gene transfer (LGT) between bacteria. Using our gene profile, we defined and identified a MLG as a group of genes that co-exists among different individual samples and has a consistent abundance level and taxonomic assignment (Supplementary Methods).

To assess the reliability of our MLG identifying method, we first constructed a subset of bacterial genes from the updated metagenome gene catalogue ($n = 130,605$) that were independently derived from 50 known gut bacterial species (Supplementary Methods). We used a threshold for the minimum gene number for a MLG of 100, above which all 50 bacterial species could be identified with an average genome coverage of 83.0% and with an accuracy in the taxonomic classification of genes in the constructed subset of 99.8% (Supplementary Fig. 8 and Supplementary Table 8).

We identified 47 MLGs in the T2D-associated gene markers, which covered 84.4% of these markers (Supplementary Table 9). Of these, 17 MLGs could be assigned to known bacterial species on the basis of strong alignment sequence similarity with sequenced bacterial genomes at the nucleotide level (Table 1). Using the taxonomic characterization from these MLGs, we found that almost all of the MLGs enriched in the control samples were from various butyrate-producing bacteria, including *Clostridiales* sp. SS3/4, *Eubacterium rectale*, *Faecalibacterium prausnitzii*, *Roseburia intestinalis* and *Roseburia inulinivorans*. By contrast, most of T2D-enriched MLGs were from opportunistic pathogens, such as *Bacteroides caccae*, *Clostridium hathewayi*, *Clostridium ramosum*, *Clostridium symbiosum*, *Eggerthella lenta* and *Escherichia coli*, which have previously been reported to cause or underlie human infections such as bacteraemia

and intra-abdominal infections^{22–25}. Of interest, the known mucin-degrading species *Akkermansia muciniphila* and sulphate-reducing species *Desulfovibrio* sp. 3_1_syn3 were also enriched in T2D samples. The MLGs that were of unknown species origin will be of interest for isolation and analysis in future studies to obtain information on their relevant taxonomy.

A co-occurrence network on these MLGs was generated to assess potential relationships between the T2D-associated gut bacteria (Fig. 2a and Supplementary Methods). In this result, some types of butyrate-producers, from clostridial cluster XIVa and IV, showed a positive correlation with one another and were negatively correlated with a group of the T2D-enriched bacteria from *Clostridium*, which may indicate an antagonistic relationship between these different clostridial clusters. Another interesting finding was the presence of a small MLG from *Haemophilus parainfluenzae*, which is not a butyrate-producer but was significantly enriched in the control samples, even in an independent analysis comparing the coverage of its sequenced bacterial genome (the highest genome coverage in all samples was 94.5%; $P < 0.001$ between case and control groups, Student's *t*-test). In the co-occurrence network, this MLG was clearly separate from the cluster of butyrate producers, and may have an unknown antagonistic relationship with a T2D-enriched bacterium that is unknown but appears closely related to the *Subdoligranulum* genus. These data presented various patterns indicating relationships between the T2D-associated gut bacteria and suggested it may be important to determine, in a case-by-case manner, the different roles gut bacteria may have in maintaining or interacting with their environment.

Functional characterization related to T2D

Using the T2D-associated KEGG orthologues and eggNOG orthologue groups markers, we assessed the potential microbial functional roles in the gut microbiota of T2D patients. In general T2D-enriched markers were typically involved in the KEGG categories of membrane transport ($P < 0.001$, Fisher's exact test). This result is consistent with

Table 1 | The list of T2D-associated MLGs that could be assigned to previously known phylotypes

MLG ID	No. of genes	P values*		Odds ratios (95% CI)†	Taxonomy assignment (level)	Percentage similarity‡
		Stage I	Stage II			
T2D-enriched						
T2D-154	337	0.0014	2.54×10^{-4}	1.52 (1.05, 2.19)	<i>Akkermansia muciniphila</i>	98.2
T2D-140	148	3.97×10^{-4}	0.0029	1.50 (1.15, 1.97)	<i>Bacteroides intestinalis</i>	98.2
T2D-139	3,386	0.0013	2.11×10^{-4}	1.66 (1.26, 2.20)	<i>Bacteroides</i> sp. 20_3	99.3
T2D-11	5,113	4.16×10^{-8}	7.58×10^{-5}	5.89 (1.39, 25.0)	<i>Clostridium bolteae</i>	99.4
T2D-5	2,378	4.21×10^{-5}	1.97×10^{-6}	23.1 (2.08, 257)	<i>Clostridium hathewayi</i>	99.3
T2D-80	2,381	1.30×10^{-4}	1.41×10^{-5}	1.68 (0.97, 2.89)	<i>Clostridium ramosum</i>	99.8
T2D-57	821	4.00×10^{-7}	2.21×10^{-5}	2.62 (1.14, 6.03)	<i>Clostridium</i> sp. HGF2	99.6
T2D-15	2,492	4.74×10^{-5}	2.97×10^{-4}	1.13 (0.88, 1.44)	<i>Clostridium symbiosum</i>	99.6
T2D-1	949	6.01×10^{-4}	0.0036	1.41 (0.93, 2.13)	<i>Desulfovibrio</i> sp. 3_1_syn3	98.0
T2D-7	1,056	6.01×10^{-4}	2.80×10^{-4}	1.57 (0.95, 2.58)	<i>Eggerthella lenta</i>	99.6
T2D-137	425	6.71×10^{-7}	0.0012	1.72 (1.16, 2.57)	<i>Escherichia coli</i>	99.0
T2D-165	131	0.0096	0.0017	1.46 (1.07, 1.99)	<i>Alistipes</i> (genus)	99.5§
T2D-12	364	4.52×10^{-6}	8.04×10^{-8}	2.22 (1.12, 4.40)	<i>Clostridium</i> (genus)	91.0
T2D-8	5,272	7.08×10^{-10}	9.95×10^{-6}	1.12 (0.86, 1.45)	<i>Clostridium</i> (genus)	88.8
T2D-93	1,590	2.01×10^{-4}	0.0020	1.84 (1.03, 3.29)	<i>Parabacteroides</i> (genus)	80.5§
T2D-62	2,584	7.63×10^{-6}	6.88×10^{-4}	2.41 (1.43, 4.08)	<i>Subdoligranulum</i> (genus)	98.7§
T2D-2	2,430	3.14×10^{-5}	0.0019	4.06 (1.28, 12.9)	<i>Lachnospiraceae</i> (family)	97.3§
Control-enriched						
Con-107	1,677	1.12×10^{-7}	0.0018	1.44 (1.13, 1.84)	<i>Clostridiales</i> sp. SS3/4	98.0
Con-112	232	0.0064	1.99×10^{-4}	1.51 (1.13, 2.03)	<i>Eubacterium rectale</i>	97.6
Con-129	1,440	0.0033	0.0010	1.55 (1.19, 2.00)	<i>Faecalibacterium prausnitzii</i>	98.2
Con-166	273	3.80×10^{-5}	1.94×10^{-4}	1.25 (0.93, 1.69)	<i>Haemophilus parainfluenzae</i>	94.8
Con-121	3,507	6.11×10^{-5}	4.90×10^{-6}	3.10 (1.92, 5.03)	<i>Roseburia intestinalis</i>	98.9
Con-113	345	2.85×10^{-4}	9.72×10^{-4}	1.45 (1.11, 1.89)	<i>Roseburia inulinivorans</i>	98.2
Con-120	116	1.90×10^{-4}	5.41×10^{-4}	1.55 (1.17, 2.06)	<i>Eubacterium</i> (genus)	89.0
Con-130	670	0.0134	0.0018	1.59 (1.21, 2.08)	<i>Faecalibacterium</i> (genus)	89.4
Con-131	202	8.99×10^{-4}	0.0017	1.58 (1.16, 2.15)	<i>Faecalibacterium</i> (genus)	96.9
Con-133	1,555	3.43×10^{-5}	0.0015	1.52 (1.15, 2.01)	<i>Erysipelotrichaceae</i> (family)	66.9§
Con-109	378	0.0135	1.67×10^{-4}	1.41 (1.09, 1.83)	<i>Clostridiales</i> (order)	87.0

* The stage I P value was calculated after adjustment for population structures, stage II P value was one-side.

† Calculated by logistic model.

‡ Similarity at nucleic acid level or, when marked with § at the protein level.

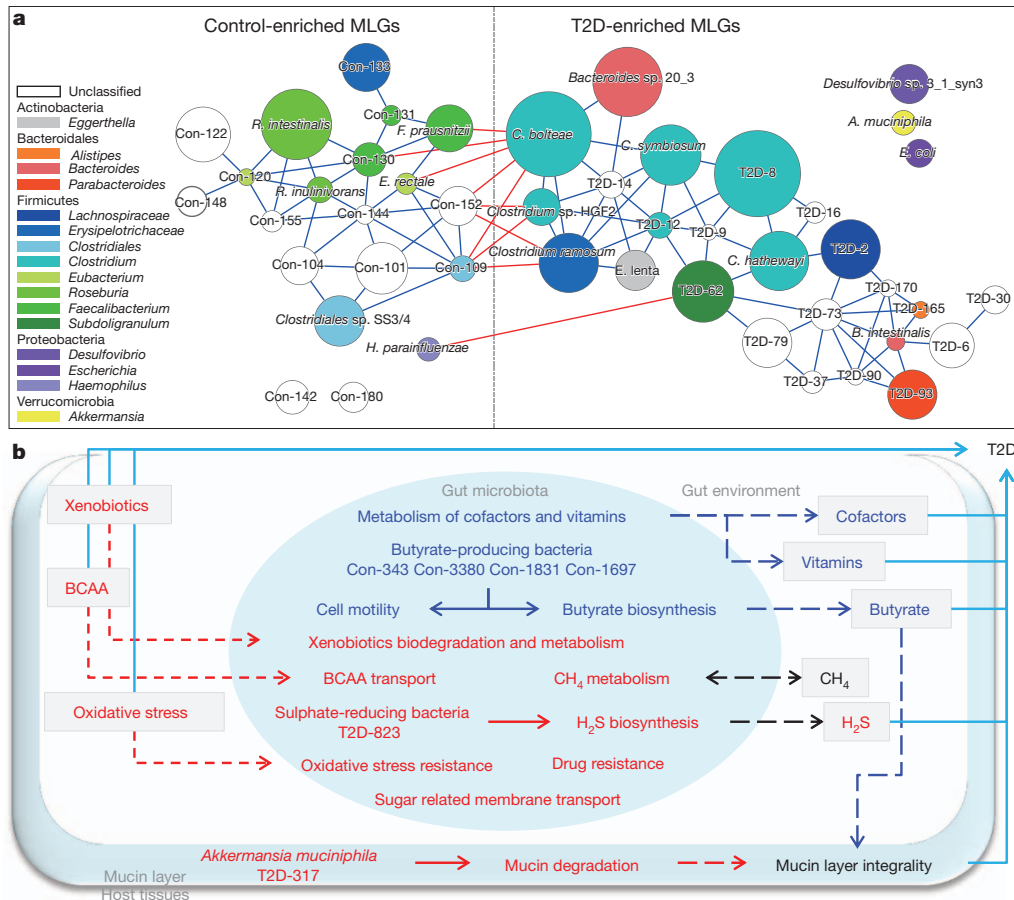


Figure 2 | Taxonomic and functional characterization of gut microbiota in T2D. **a**, A co-occurrence network was deduced from 47 MLGs that were identified from 52,484 gene markers. Nodes depict MLGs with their ID displayed in the centre. The size of the nodes indicates gene number within the MLG. The colour of the nodes indicates their taxonomic assignment. Connecting lines represent Spearman correlation coefficient values above 0.4

the previous findings in studies of inflammatory bowel disease and obese patients²⁶. By contrast, control-enriched markers were frequently involved in cell motility and metabolism of cofactors and vitamins ($P < 0.002$; Supplementary Fig. 9).

At the module or pathway level, the gut microbiota of T2D patients was functionally characterized with our T2D-associated markers and showed enrichment in membrane transport of sugars, branched-chain amino acid (BCAA) transport, methane metabolism, xenobiotics degradation and metabolism, and sulphate reduction. By contrast, there was a decrease in the level of bacterial chemotaxis, flagellar assembly, butyrate biosynthesis and metabolism of cofactors and vitamins (Fig. 2b and Supplementary Table 10; see Supplementary Fig. 10 for the detailed information on butyrate-CoA transferase). Some important functions, including butyrate biosynthesis and sulphate reduction, coincided with the T2D-associated bacteria identified in the MLG analysis. The butyrate-producing bacteria seemed to be the primary contributors to the cell motility functions (Supplementary Table 11), potentially indicating some functional enrichment might be related to the presence of specific species enrichment.

We found that seven of the T2D-enriched KEGG orthologues markers were related to oxidative stress resistance, including catalase (K03781), peroxiredoxin (K03386), Mn-containing catalase (K07217), glutathione reductase (NADPH) (K00383), nitric oxide reductase (K02448), putative iron-dependent peroxidase (K07223), and cytochrome *c* peroxidase (K00428), but none of the identified control-enriched KEGG orthologues markers had similar types of function.

(blue) or below -0.4 (red). **b**, A schematic diagram showing the main functions of the gut microbes that had a predicted T2D association. Red text denotes enriched functions in T2D patients; blue text denotes depleted functions in T2D patients; black text denotes an uncertain functional role relative to T2D. The dashed line arrows point to the inference that was not detected directly but reported by previous studies.

This may indicate that the gut environment of a T2D patient is one that stimulates bacterial defence mechanisms against oxidative stress (Supplementary Table 10). Similarly, we found 14 KEGG orthologues markers related to drug resistance that were greatly enriched in T2D patients, further supporting that T2D patients may have a more hostile gut environment, and the medical histories of these patients may reflect this (Supplementary Table 10).

T2D-related dysbiosis in gut microbiota

In light of the above MGWAS result and an additional PERMANOVA²⁷ (permutational multivariate analysis of variance) analysis that clearly showed that T2D was a significant factor for explaining the variation in the examined gut microbial samples (Supplementary Table 12), we deduced that the gut microbiota in T2D patients featured dysbiosis, which is a state where the balance of the normal microbiota has been disturbed. However, the degree of this T2D-related dysbiosis was moderate, because only $3.8 \pm 0.2\%$ (mean \pm s.e.m.; $n = 344$) of the gut microbial genes (at the relative abundance level) were associated with T2D in an individual. Additionally, we did not observe a significant difference in the within-sample diversity between T2D and control groups (Fig. 3a). Specifically, the degree of gut microbiota change in T2D was not as substantial as that seen in inflammatory bowel disease (from the MetaHIT samples⁸; see Fig. 3a) or enterotypes (Supplementary Fig. 11). A similar result using the eggNOG orthologue groups profile supported the same conclusion (Supplementary Fig. 12).

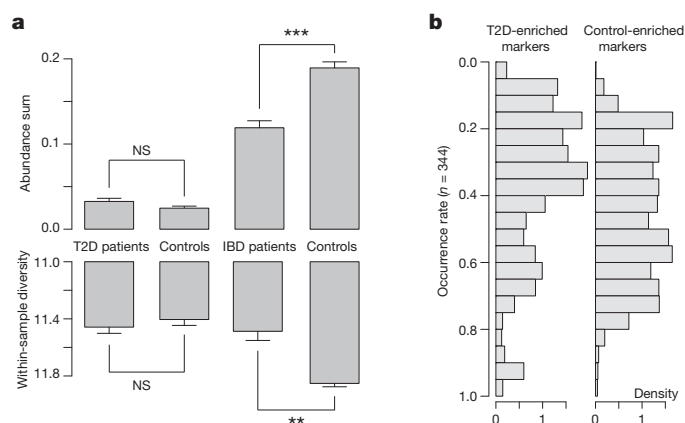


Figure 3 | Gut microbiota of T2D patients show a moderate degree of dysbiosis. **a**, An ecological comparison between T2D patients ($n = 170$) and control ($n = 174$) in all samples, as well as inflammatory bowel disease (IBD) patients ($n = 25$) and control ($n = 99$) from published MetaHIT samples⁸. The upward bars denote the gross relative abundance of the T2D-associated gene markers for each sample and the same value computed on the inflammatory-bowel-disease-associated gene markers (see Supplementary Methods). The downward bars denote the within-sample diversity (calculated using the Shannon index) in each group. For an individual sample, a lower proportion of gut microbiota was implicated in T2D disease and there was no significant difference in the within-sample diversity between the T2D patients and control as compared with the distinct difference seen in the inflammatory bowel disease analysis. ** $P < 0.01$; *** $P < 0.001$ (Student's t -test); NS, not significant; and the error bar denote standard error. **b**, A density histogram showing a comparison of the occurrence rate distribution between T2D-enriched gene markers and control-enriched gene markers in all samples ($n = 344$). The threshold of mapped read number for gene identification is ≥ 2 .

To characterize ecologically the gut bacteria involved in the T2D-related dysbiosis, we compared, in all individual samples, the distribution of the occurrence rate of both T2D-associated gene and function markers, and these showed the same pattern, which was that the control-enriched markers had a higher occurrence rate on average than the T2D-enriched markers (Fig. 3b and Supplementary Figs 13–15). This may be because the beneficial bacteria lost in the T2D gut were universally present, whereas some of the harmful bacteria that appeared in the T2D gut were diverse, and thus had less overall abundance within the human population.

Gut-microbiota-based T2D classification

To exploit the potential ability of T2D classification by gut microbiota, we developed a T2D classifier system based on the 50 gene markers that we defined as an optimal gene set by a minimum redundancy–maximum relevance (mRMR) feature selection method (Supplementary Fig. 16 and Supplementary Table 13). For intuitive evaluation of the risk of T2D disease based on these 50 gut microbial gene markers, we computed a T2D index (Supplementary Methods), which correlated well with the ratio of T2D patients in our population (Fig. 4a), and the area under the receiver operating characteristic (ROC) curve was 0.81 (95% confidence interval 0.76–0.85) (Fig. 4b), indicating the gut-microbiota-based T2D index could be used to classify T2D individuals accurately.

We validated the discriminatory power of our T2D classifier using an independent study group: 11 T2D patients and 12 non-diabetic controls. In this assessment analysis, the top eight samples with the highest T2D index were all T2D patients (Fig. 4c and Supplementary Table 14); the average T2D index between case and control was significantly different ($P = 0.004$, Student's t -test). Overall, our cross-sectional study in overt T2D indicated that it would be worthwhile to test more extensively gut-microbiota-based classifiers in future longitudinal studies for their ability to identify subsets of the population that are at high risk for progressing to clinically defined T2D.

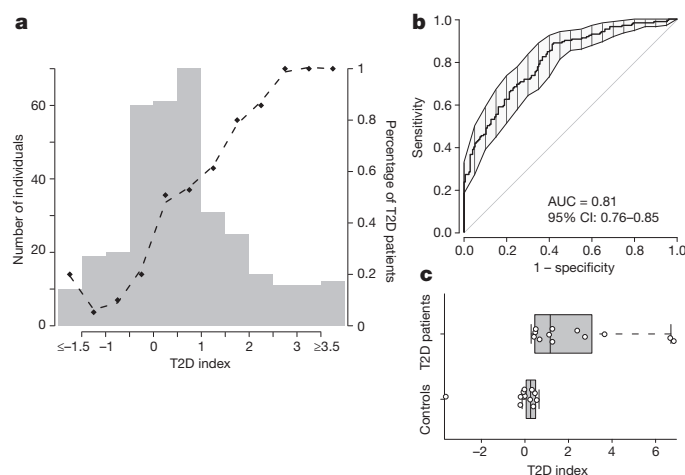


Figure 4 | A trial classification of T2D using gut microbial gene markers.

a, A classifier to identify T2D individuals was constructed using 50 gene markers selected by mRMR, and then, for each individual, a T2D index was calculated to evaluate the risk of T2D. The histogram shows the distribution of T2D indices for all individuals, in which values less than -1.5 and values greater than 3.5 were grouped. For each bin, the black dots show the proportion of T2D patients in the population of that bin (y axis on the right). **b**, The area under the ROC curve (AUC) of gut-microbiota-based T2D classification. The black bars denote the 95% confidence interval (CI) and the area between the two outside curves represents the 95% CI shape. **c**, The T2D index was computed for an additional 11 Chinese T2D samples and 12 non-diabetic controls. The box depicts the interquartile range (IQR) between the first and third quartiles (25th and 75th percentiles, respectively) and the line inside denotes the median, whereas the points represent the T2D index in each sample.

Discussion

T2D is a heterogeneous and multifactorial disease, influenced by a number of different genetic and environmental factors. By applying the standard two-stage GWAS strategy to design and carry out a MGWAS to identify disease-associated metagenomic markers, the present study highlights how the gut microbial composition, traditionally considered to be factors of environmental origin¹², differs between T2D patients and non-diabetic control subjects in a Chinese population.

We first established an updated human microbial gene reference set, adding information from both a new ethnicity and from T2D patients, which will be a useful resource for future metagenomic analyses. We also developed the concept of a MLG, which provided various types of taxonomic information from whole-genome shotgun data, including bacterial species-specific regions on a chromosome, and mobile genetic elements, such as plasmids and bacteriophages. Thus, a MLG can provide metagenomic species-level information even for unknown species, instead of requiring traditional taxonomic classification approaches based on sequence composition or similarity^{28,29}. The use of species-level information allows assessment of the relationships between the T2D-associated bacteria. For example, we identified what appears to be an antagonistic relationship between beneficial bacteria and harmful bacteria, highlighted by the large populations of clostridial clusters. These species-level analyses also showed various patterns: for example, the MLG from *Haemophilus parainfluenzae* in the control samples could be inferred, under these circumstances, to be beneficial; however, on the basis of relationship patterns, it was quite distinct from the other inferred beneficial bacteria, indicating that *H. parainfluenzae* may have a different type of impact in this specific biological context (Fig. 2a).

Our findings indicated that T2D patients had only a moderate degree gut bacterial dysbiosis; however, functional annotation analyses indicated a decline in butyrate-producing bacteria, which may be metabolically beneficial, and an increase in several opportunistic

pathogens. Importantly, the abundance of these categories of opportunistic pathogens seemed to be quite diverse among our Chinese study participants. Such changes in the intestinal bacteria composition have recently been reported for colorectal cancer patients³⁰ and ageing population³¹. Thus, a general picture is emerging where butyrate-producing bacteria seem to have a protective role against several types of diseases. Additionally, our finding of a general dysbiosis in T2D patients raises the possibility that there is a 'functional dysbiosis', rather than there being a specific microbial species that has a direct association with T2D pathophysiology. Furthermore, given that other intestinal diseases show a loss of butyrate-producing bacteria with a commensurate increase in opportunistic pathogens, it is possible that dysbiosis that results in a disordered, rather than directional, alteration of gut microbial composition may itself have a role in increasing the susceptibility to a variety of diseases.

Our analysis of bacterial gene functions indicating there was an increase in functions relating to gut oxidative stress response is also of interest, given that previous studies have shown that a high oxidative stress level is related to a predisposition for diabetic complications³². Finally, our findings that gut metagenomic markers are able to differentiate between T2D cases and controls with a higher level of specificity than similar analyses based on human genome variation³³ raises the possibility for a mode of monitoring gut health and a complementary approach for risk assessment of this common disorder.

METHODS SUMMARY

Sample collection and DNA extraction. Faecal samples were obtained from 368 volunteers (345 samples for MGWAS and 23 additional samples for T2D classification) after signing an informed consent form. The sampling procedure was approved by the Ethical Committee for Clinical Research from the Peking University Shenzhen Hospital, Shenzhen Second People's Hospital and Medical Research Center of Guangdong General Hospital. The individuals had not received any antibiotic treatment within 2 months before sample collection. The samples were frozen immediately and underwent DNA extraction using standard methods³⁴.

Sequencing and data processing. Illumina GAIIx and HiSeq 2000 were used to sequence the samples. We constructed a paired-end library with insert size of ~350 base pairs for every sample. Adaptor contamination and low-quality reads were discarded from the raw reads, and the remaining reads were filtered to eliminate human host DNA based on the human genome reference (hg18).

Full Methods and associated references are available in the Supplementary Information.

Received 30 August 2011; accepted 27 July 2012.

Published online 26 September 2012.

- Wellen, K. E. & Hotamisligil, G. S. Inflammation, stress, and diabetes. *J. Clin. Invest.* **115**, 1111–1119 (2005).
- Risérus, U., Willett, W. C. & Hu, F. B. Dietary fats and prevention of type 2 diabetes. *Prog. Lipid Res.* **48**, 44–51 (2009).
- The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
- Scott, L. J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
- Musso, G., Gambino, R. & Cassader, M. Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. *Annu. Rev. Med.* **62**, 361–380 (2011).
- Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638 (2005).
- Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
- Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
- The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
- The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
- Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228–231 (2010).
- Bäckhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl Acad. Sci. USA* **101**, 15718–15723 (2004).
- Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102**, 11070–11075 (2005).

- Zhang, H. *et al.* Human gut microbiota in obesity and after gastric bypass. *Proc. Natl Acad. Sci. USA* **106**, 2365–2370 (2009).
- Bäckhed, F., Manchester, J. K., Semenkovich, C. F. & Gordon, J. I. Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proc. Natl Acad. Sci. USA* **104**, 979–984 (2007).
- Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
- Manichanh, C. *et al.* Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55**, 205–211 (2006).
- Joossens, M. *et al.* Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* **60**, 631–637 (2011).
- Larsen, N. *et al.* Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS ONE* **5**, e9085 (2010).
- Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
- Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* **38**, 904–909 (2006).
- Woo, P. C. Y. *et al.* Bacteremia due to *Clostridium hathewayi* in a patient with acute appendicitis. *J. Clin. Microbiol.* **42**, 5947–5949 (2004).
- Elsayed, S. & Zhang, K. Bacteremia caused by *Clostridium symbiosum*. *J. Clin. Microbiol.* **42**, 4390–4392 (2004).
- McClean, K. L., Sheehan, G. J. & Harding, G. K. Intraabdominal infection: a review. *Clin. Inf. Dis.* **19**, 100–116 (1994).
- Brook, I. Clostridial infection in children. *J. Med. Microbiol.* **42**, 78–82 (1995).
- Greenblum, S., Turnbaugh, P. J. & Borenstein, E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc. Natl Acad. Sci. USA* **109**, 594–599 (2012).
- McArdle, B. H. & Anderson, M. J. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **82**, 290–297 (2001).
- Yang, B. *et al.* Unsupervised binning of environmental genomic fragments based on an error robust selection of l-mers. *BMC Bioinformatics* **11** (suppl. 2), S5 (2010).
- Krause, L. *et al.* Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Res.* **36**, 2230–2239 (2008).
- Wang, T. *et al.* Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J.* **6**, 320–329 (2012).
- Biagi, E. *et al.* Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS ONE* **5**, e10667 (2010).
- Kashyap, P. & Farrugia, G. Oxidative stress: key player in gastrointestinal complications of diabetes. *Neurogastroenterol. Motil.* **23**, 111–114 (2011).
- Lyssenko, V. *et al.* Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N. Engl. J. Med.* **359**, 2220–2232 (2008).
- Godon, J. J., Zumstein, E., Dabert, P., Habouzit, F. & Moletta, R. Molecular microbial diversity of an anaerobic digester as determined by small-subunit rDNA sequence analysis. *Appl. Environ. Microbiol.* **63**, 2802–2813 (1997).
- Li, S. *et al.* Type 2 diabetes gut metagenome (microbiome) data from 368 Chinese samples. *GigaScience* <http://dx.doi.org/10.5524/100036> (2012).
- Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Goodman for editing the manuscript and providing comments. This research was supported by the Ministry of Science and Technology of China, 863 program (2012AA02A201), the National Natural Science Foundation of China (30890032, 30725008, 30811130531, 31161130357), the Shenzhen Municipal Government of China (ZXC200903240080A, BGI20100001, CXB201108250096A, CXB201108250098A), the Danish Strategic Research Council grant (2106-07-0021), the Ole Rømer grant from Danish Natural Science Research Council, the Solexa project (272-07-0196), and the European Commission FP7 grant HEALTH-F4-2007-201052. The Lundbeck Foundation Centre for Applied Medical Genomics in Personalised Disease Prediction, Prevention and Care (LuCamp, www.lucamp.org). The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (<http://www.metabol.ku.dk>). We are also indebted to many additional faculty and staff of BGI-Shenzhen who contributed to this work.

Author Contributions The project idea was conceived and the project was designed by Ju.W., K.K., O.P., R.N. and S.D.E.; J.Q., Y.L., Sh.L. and Ju.W. managed the project. F.Z., Z.C., R.X., Su.L., L.H., D.L., P.W., Y.D., X.S., Z.L., A.T., S.Z., M.W., Q.F. and T.H. performed sample collection and clinical study. Wen.Z., M.G., J.Y., Y.Z. and W.X. performed DNA experiments. Ju.W., K.K., O.P., R.N., S.D.E., J.Q., Y.L., Sh.L. and J.Z. designed the analysis. J.Q., Y.L., Sh.L., J.Z., Su.L., Y.G., Y.P., D.S., X.L., W.C., D.Z., Y.Q., M.Z., Z.Z., J.J., G.S., J.L., J.R., S.O., H.C. and W.W. performed the data analysis. J.Q., Sh.L., J.Z., Y.G., Y.P., M.A., E.L., P.R., N.P. and J.-M.B. worked on metagenomic linkage group method. J.Q., D.S., Su.L., Y.Q., J.R., G.F. and S.O. did the functional annotation analyses. J.Q., Sh.L., D.S., J.Z., Y.P. and Y.L. wrote the paper. Ju.W., O.P., K.K., R.N., S.D.E., Ji.W., H.Y., So.L., Wei.Z. and R.Y. revised the paper.

Author Information The raw Illumina read data of all 368 samples has been deposited in the NCBI Sequence Read Archive under accession numbers SRA045646 and SRA050230. The assembly data, updated metagenome gene catalogue, annotation information, and MGLs are published in the *GigaScience* database, *GigaDB*³⁵. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Ju.W. (wangj@genomics.org.cn).

Comprehensive molecular portraits of human breast tumours

The Cancer Genome Atlas Network*

We analysed primary breast cancers by genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, microRNA sequencing and reverse-phase protein arrays. Our ability to integrate information across platforms provided key insights into previously defined gene expression subtypes and demonstrated the existence of four main breast cancer classes when combining data from five platforms, each of which shows significant molecular heterogeneity. Somatic mutations in only three genes (*TP53*, *PIK3CA* and *GATA3*) occurred at >10% incidence across all breast cancers; however, there were numerous subtype-associated and novel gene mutations including the enrichment of specific mutations in *GATA3*, *PIK3CA* and *MAP3K1* with the luminal A subtype. We identified two novel protein-expression-defined subgroups, possibly produced by stromal/microenvironmental elements, and integrated analyses identified specific signalling pathways dominant in each molecular subtype including a HER2/phosphorylated HER2/EGFR/phosphorylated EGFR signature within the HER2-enriched expression subtype. Comparison of basal-like breast tumours with high-grade serous ovarian tumours showed many molecular commonalities, indicating a related aetiology and similar therapeutic opportunities. The biological finding of the four main breast cancer subtypes caused by different subsets of genetic and epigenetic abnormalities raises the hypothesis that much of the clinically observable plasticity and heterogeneity occurs within, and not across, these major biological subtypes of breast cancer.

Breast cancer is one of the most common cancers with greater than 1,300,000 cases and 450,000 deaths each year worldwide. Clinically, this heterogeneous disease is categorized into three basic therapeutic groups. The oestrogen receptor (ER) positive group is the most numerous and diverse, with several genomic tests to assist in predicting outcomes for ER⁺ patients receiving endocrine therapy^{1,2}. The *HER2* (also called *ERBB2*) amplified group³ is a great clinical success because of effective therapeutic targeting of *HER2*, which has led to intense efforts to characterize other DNA copy number aberrations^{4,5}. Triple-negative breast cancers (TNBCs, lacking expression of ER, progesterone receptor (PR) and *HER2*), also known as basal-like breast cancers⁶, are a group with only chemotherapy options, and have an increased incidence in patients with germline *BRCA1* mutations^{7,8} or of African ancestry⁹.

Most molecular studies of breast cancer have focused on just one or two high information content platforms, most frequently mRNA expression profiling or DNA copy number analysis, and more recently massively parallel sequencing^{10–12}. Supervised clustering of mRNA expression data has reproducibly established that breast cancers encompass several distinct disease entities, often referred to as the intrinsic subtypes of breast cancer^{13,14}. The recent development of additional high information content assays focused on abnormalities in DNA methylation, microRNA (miRNA) expression and protein expression, provide further opportunities to characterize more completely the molecular architecture of breast cancer. In this study, a diverse set of breast tumours were assayed using six different technology platforms. Individual platform and integrated pathway analyses identified many subtype-specific mutations and copy number changes that identify therapeutically tractable genomic aberrations and other events driving tumour biology.

Samples and clinical data

Tumour and germline DNA samples were obtained from 825 patients. Different subsets of patients were assayed on each platform:

466 tumours from 463 patients had data available on five platforms including Agilent mRNA expression microarrays ($n = 547$), Illumina Infinium DNA methylation chips ($n = 802$), Affymetrix 6.0 single nucleotide polymorphism (SNP) arrays ($n = 773$), miRNA sequencing ($n = 697$), and whole-exome sequencing ($n = 507$); in addition, 348 of the 466 samples also had reverse-phase protein array (RPPA) data ($n = 403$). Owing to the short median overall follow up (17 months) and the small number of overall survival events (93 out of 818), survival analyses will be presented in a later publication. Demographic and clinical characteristics are presented in Supplementary Table 1.

Significantly mutated genes in breast cancer

Overall, 510 tumours from 507 patients were subjected to whole-exome sequencing, identifying 30,626 somatic mutations comprised of 28,319 point mutations, 4 dinucleotide mutations, and 2,302 insertions/deletions (indels) (ranging from 1 to 53 nucleotides). The point mutations included 6,486 silent, 19,045 missense, 1,437 nonsense, 26 read-through, 506 splice-site mutations, and 819 mutations in RNA genes. Comparison to COSMIC and OMIM databases identified 619 mutations across 177 previously reported cancer genes. Of 19,045 missense mutations, 9,484 were predicted to have a high probability of being deleterious by Condel¹⁵. The MuSiC package¹⁶, which determines the significance of the observed mutation rate of each gene based on the background mutation rate, identified 35 significantly mutated genes (excluding LOC or Ensembl gene IDs) by at least two tests (convolution and likelihood ratio tests) with false discovery rate (FDR) <5% (Supplementary Table 2).

In addition to identifying nearly all genes previously implicated in breast cancer (*PIK3CA*, *PTEN*, *AKT1*, *TP53*, *GATA3*, *CDH1*, *RBI*, *MLL3*, *MAP3K1* and *CDKN1B*), a number of novel significantly mutated genes were identified including *TBX3*, *RUNX1*, *CBFB*, *AFF2*, *PIK3R1*, *PTPN22*, *PTPRD*, *NF1*, *SF3B1* and *CCND3*. *TBX3*, which is mutated in ulnar-mammary syndrome and involved in mammary gland development¹⁷, harboured 13 mutations (8 frame-shift indels,

*A list of participants and their affiliations appears at the end of the paper.

1 in-frame deletion, 1 nonsense, and 3 missense), suggesting a loss of function. Additionally, 2 mutations were found in *TBX4* and 1 mutation in *TBX5*, which are genes involved in Holt–Oram syndrome¹⁸. Two other transcription factors, *CTCF* and *FOXA1*, were at or near significance harbouring 13 and 8 mutations, respectively. *RUNX1* and *CBFB*, both rearranged in acute myeloid leukaemia and interfering with haematopoietic differentiation, harboured 19 and 9 mutations, respectively. *PIK3R1* contained 14 mutations, most of which clustered in the PIK3CA interaction domain similar to previously identified mutations in glioma¹⁹ and endometrial cancer²⁰. We also observed a statistically significant exclusion pattern among *PIK3R1*, *PIK3CA*, *PTEN* and *AKT1* mutations ($P = 0.025$). Mutation of splicing factor *SF3B1*, previously described in myelodysplastic syndromes²¹ and chronic lymphocytic leukaemia²², was significant with 15 non-silent mutations, of which 4 were a recurrent K700E substitution. Two protein tyrosine phosphatases (*PTPN22* and *PTPRD*) were also significantly mutated; frequent deletion/mutation of *PTPRD* is observed in lung adenocarcinoma²³.

Mutations and mRNA-expression subtype associations

We analysed the somatic mutation spectrum within the context of the four mRNA-expression subtypes, excluding the normal-like group owing to small numbers ($n = 8$) (Fig. 1). Several significantly mutated genes showed mRNA-subtype-specific (Supplementary Figs 1–3) and clinical-subtype-specific patterns of mutation (Supplementary Table 2). Significantly mutated genes were considerably more diverse and recurrent within luminal A and luminal B tumours than within basal-like and HER2-enriched (HER2E) subtypes; however, the overall mutation rate was lowest in luminal A subtype and highest in the basal-like and HER2E subtypes. The luminal A subtype harboured the most significantly mutated genes, with the most frequent being *PIK3CA* (45%), followed by *MAP3K1*, *GATA3*, *TP53*, *CDH1* and *MAP2K4*. Twelve per cent of luminal A tumours contained likely inactivating mutations in *MAP3K1* and *MAP2K4*, which represent two contiguous

steps in the p38–JNK1 stress kinase pathway²⁴. Luminal B cancers exhibited a diversity of significantly mutated genes, with *TP53* and *PIK3CA* (29% each) being the most frequent. The luminal tumour subtypes markedly contrasted with basal-like cancers where *TP53* mutations occurred in 80% of cases and the majority of the luminal significantly mutated gene repertoire, except *PIK3CA* (9%), were absent or near absent. The HER2E subtype, which has frequent *HER2* amplification (80%), had a hybrid pattern with a high frequency of *TP53* (72%) and *PIK3CA* (39%) mutations and a much lower frequency of other significantly mutated genes including *PIK3R1* (4%).

Intrinsic mRNA subtypes differed not only by mutation frequencies but also by mutation type. Most notably, *TP53* mutations in basal-like tumours were mostly nonsense and frame shift, whereas missense mutations predominated in luminal A and B tumours (Supplementary Fig. 1). Fifty-eight somatic *GATA3* mutations, some of which were previously described²⁵, were detected including a hotspot 2-base-pair deletion within intron 4 only in the luminal A subtype (13 out of 13 mutants) (Supplementary Fig. 2). In contrast, 7 out of 9 frame-shift mutations in exon 5 (DNA binding domain) occurred in luminal B cancers. *PIK3CA* mutation frequency and spectrum also varied by mRNA subtype (Supplementary Fig. 3); the recurrent *PIK3CA* E545K mutation was present almost exclusively within luminal A (25 out of 27) tumours. *CDH1* mutations were common (30 out of 36) within the lobular histological subtype and corresponded with lower *CDH1* mRNA (Supplementary Fig. 4) and protein expression. Finally, we identified 4 out of 8 somatic variants in *HER2* within lobular cancers, three of which were within the tyrosine kinase domain.

We performed analyses on a selected set of genes²⁶ using the normal tissue DNA data and detected a number of germline predisposing variants. These analyses identified 47 out of 507 patients with deleterious germline variants, representing nine different genes (*ATM*, *BRCA1*, *BRCA2*, *BRIP1*, *CHEK2*, *NBN*, *PTEN*, *RAD51C* and *TP53*; Supplementary Table 3), supporting the hypothesis that ~10% of sporadic breast cancers may have a strong germline contribution.

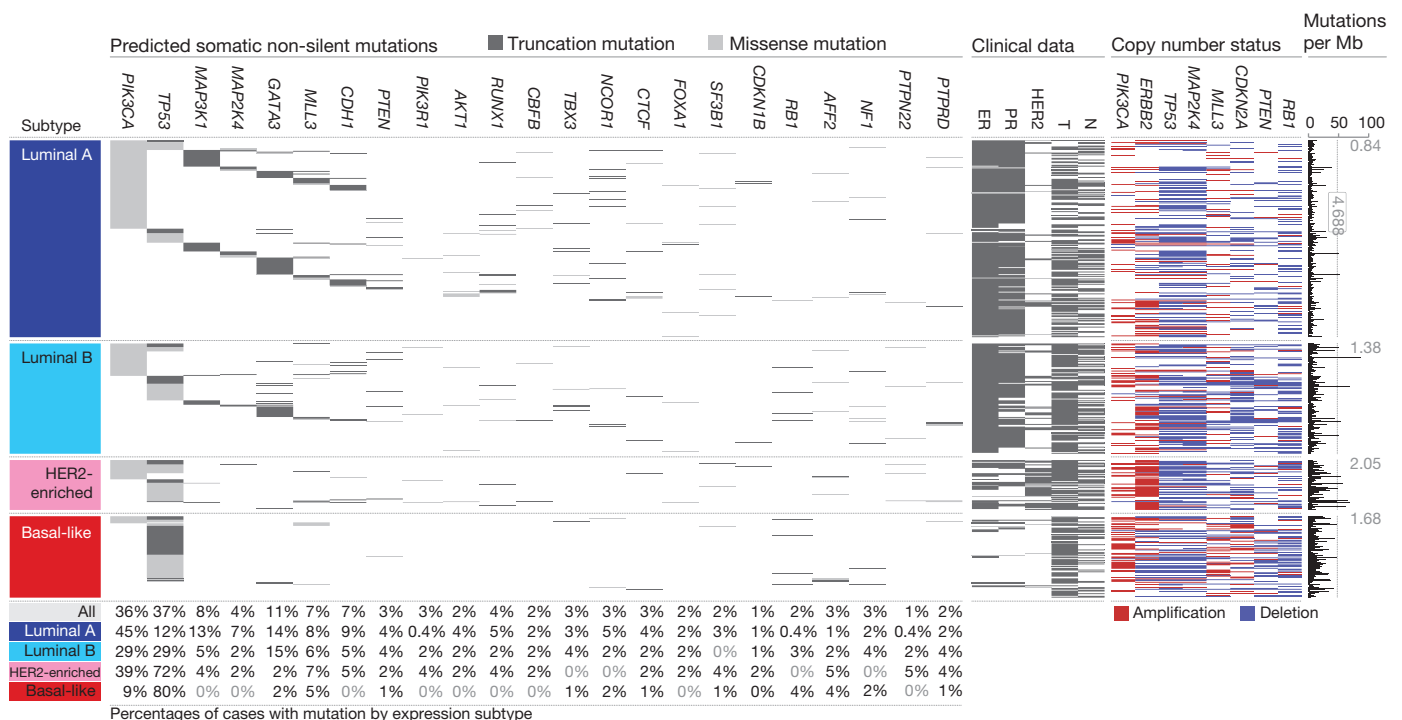


Figure 1 | Significantly mutated genes and correlations with genomic and clinical features. Tumour samples are grouped by mRNA subtype: luminal A ($n = 225$), luminal B ($n = 126$), HER2E ($n = 57$) and basal-like ($n = 93$). The left panel shows non-silent somatic mutation patterns and frequencies for significantly mutated genes. The middle panel shows clinical features: dark grey, positive or T2–4; white, negative or T1; light grey, N/A or equivocal.

N, node status; T, tumour size. The right panel shows significantly mutated genes with frequent copy number amplifications (red) or deletions (blue). The far-right panel shows non-silent mutation rate per tumour (mutations per megabase, adjusted for coverage). The average mutation rate for each expression subtype is indicated. Hypermutated: mutation rates >3 s.d. above the mean (>4.688, indicated by grey line).

These data confirmed the association between the presence of germline *BRCA1* mutations and basal-like breast cancers^{7,8}.

Gene expression analyses (mRNA and miRNA)

Several approaches were used to look for structure in the mRNA expression data. We performed an unsupervised hierarchical clustering analysis of 525 tumours and 22 tumour-adjacent normal tissues using the top 3,662 variably expressed genes (Supplementary Fig. 5); SigClust analysis identified 12 classes (5 classes with >9 samples per class). We performed a semi-supervised hierarchical cluster analysis using a previously published 'intrinsic gene list'¹⁴, which identified 13 classes (9 classes with >9 samples per class) (Supplementary Fig. 6). We also classified each sample using the 50-gene PAM50 model¹⁴ (Supplementary Fig. 5). High concordance was observed between all three analyses; therefore, we used the PAM50-defined subtype predictor as a common classification metric. There were only eight normal-like and eight claudin-low tumours²⁷, thus we did not perform focussed analyses on these two subtypes.

MicroRNA expression levels were assayed via Illumina sequencing, using 1,222 miRBase²⁸ v16 mature and star strands as the reference database of miRNA transcripts/genes. Seven subtypes were identified by consensus non-negative matrix factorization (NMF) clustering using an abundance matrix containing the 25% most variable miRNAs (306 transcripts/genes or MIMATs (miRNA IDs)). These subtypes correlated with mRNA subtypes, ER, PR and HER2 clinical status (Supplementary Fig. 7). Of note, miRNA groups 4 and 5 showed high overlap with the basal-like mRNA subtype and contained many *TP53* mutations. The remaining miRNA groups (1–3, 6 and 7) were composed of a mixture of luminal A, luminal B and HER2E with little correlation with the PAM50 defined subtypes. With the exception of *TP53*—which showed a strong positive correlation—and *PIK3CA* and *GATA3*—which showed negative associations with groups 4 and 5, respectively—there was little correlation with mutation status and miRNA subtype.

DNA methylation

Illumina Infinium DNA methylation arrays were used to assay 802 breast tumours. Data from HumanMethylation27 (HM27) and HumanMethylation450 (HM450) arrays were combined and filtered to yield a common set of 574 probes used in an unsupervised clustering analysis, which identified five distinct DNA methylation groups (Supplementary Fig. 8). Group 3 showed a hypermethylated phenotype and was significantly enriched for luminal B mRNA subtype and under-represented for *PIK3CA*, *MAP3K1* and *MAP2K4* mutations. Group 5 showed the lowest levels of DNA methylation, overlapped with the basal-like mRNA subtype, and showed a high frequency of *TP53* mutations. HER2-positive (HER2⁺) clinical status, or the HER2E mRNA subtype, had only a modest association with the methylation subtypes.

A supervised analysis of the DNA methylation and mRNA expression data was performed to compare DNA methylation group 3 ($N = 49$) versus all tumours in groups 1, 2 and 4 (excluding group 5, which consisted predominantly of basal-like tumours). This analysis identified 4,283 genes differentially methylated (3,735 higher in group 3 tumours) and 1,899 genes differentially expressed (1,232 downregulated); 490 genes were both methylated and showed lower expression in group 3 tumours (Supplementary Table 4). A DAVID (database for annotation, visualization and integrated discovery) functional annotation analysis identified 'extracellular region part' and 'Wnt signalling pathway' to be associated with this 490-gene set; the group 3 hypermethylated samples showed fewer *PIK3CA* and *MAP3K1* mutations, and lower expression of Wnt-pathway genes.

DNA copy number

A total of 773 breast tumours were assayed using Affymetrix 6.0 SNP arrays. Segmentation analysis and GISTIC were used to

identify focal amplifications/deletions and arm-level gains and losses (Supplementary Table 5). These analyses confirmed all previously reported copy number variations and highlighted a number of significantly mutated genes including focal amplification of regions containing *PIK3CA*, *EGFR*, *FOXA1* and *HER2*, as well as focal deletions of regions containing *MLL3*, *PTEN*, *RB1* and *MAP2K4* (Supplementary Fig. 9); in all cases, multiple genes were included within each altered region. Importantly, many of these copy number changes correlated with mRNA subtype including characteristic loss of 5q and gain of 10p in basal-like cancers^{5,29} and gain of 1q and/or 16q loss in luminal tumours⁴. NMF clustering of GISTIC segments identified five copy number clusters/groups that correlated with mRNA subtypes, ER, PR and HER2 clinical status, and *TP53* mutation status (Supplementary Fig. 10). In addition, this aCGH subtype classification was highly correlated with the aCGH subtypes recently defined by ref. 30 (Supplementary Fig. 11).

Reverse phase protein arrays

Quantified expression of 171 cancer-related proteins and phosphoproteins by RPPA was performed on 403 breast tumours³¹. Unsupervised hierarchical clustering analyses identified seven subtypes; one class contained too few cases for further analysis (Supplementary Fig. 12). These protein subtypes were highly concordant with the mRNA subtypes, particularly with basal-like and HER2E mRNA subtypes. Closer examination of the HER2-containing RPPA-defined subgroup showed coordinated overexpression of HER2 and EGFR with a strong concordance with phosphorylated HER2 (pY1248) and EGFR (pY992), probably from heterodimerization and cross-phosphorylation. Although there is a potential for modest cross reactivity of antibodies against these related total and phospho-proteins, the concordance of phosphorylation of HER2 and EGFR was confirmed using multiple independent antibodies.

In RPPA-defined luminal tumours, there was high protein expression of ER, PR, AR, BCL2, GATA3 and INPP4B, defining mostly luminal A cancers and a second more heterogeneous protein subgroup composed of both luminal A and luminal B cancers. Two potentially novel protein-defined subgroups were identified: reactive I consisted primarily of a subset of luminal A tumours, whereas reactive II consisted of a mixture of mRNA subtypes. These groups are termed 'reactive' because many of the characteristic proteins are probably produced by the microenvironment and/or cancer-activated fibroblasts including fibronectin, caveolin 1 and collagen VI. These two RPPA groups did not have a marked difference in the percentage tumour cell content when compared to each other, or the other protein subtypes, as assessed by SNP array analysis or pathological examination. In addition, supervised analyses of reactive I versus II groups using miRNA expression, DNA methylation, mutation, or DNA copy number data identified no significant differences between these groups, whereas similar supervised analyses using protein and mRNA expression identified many differences.

Multiplatform subtype discovery

To reveal higher-order structure in breast tumours based on multiple data types, significant clusters/subtypes from each of five platforms were analysed using a multiplatform data matrix subjected to unsupervised consensus clustering (Fig. 2). This 'cluster of clusters' (C-of-C) approach illustrated that basal-like cancers had the most distinct multiplatform signature as all the different platforms for the basal-like groups clustered together. To a great extent, the four major C-of-C subdivisions correlated well with the previously published mRNA subtypes (driven, in part, by the fact that the four intrinsic subtypes were one of the inputs). Therefore, we also performed C-of-C analysis with no mRNA data present (Supplementary Fig. 13) or with the 12 unsupervised mRNA subtypes (Supplementary Fig. 14), and in each case 4–6 groups were identified. Recent work identified ten copy-number-based subgroups in a 997 breast cancer set³⁰. We evaluated

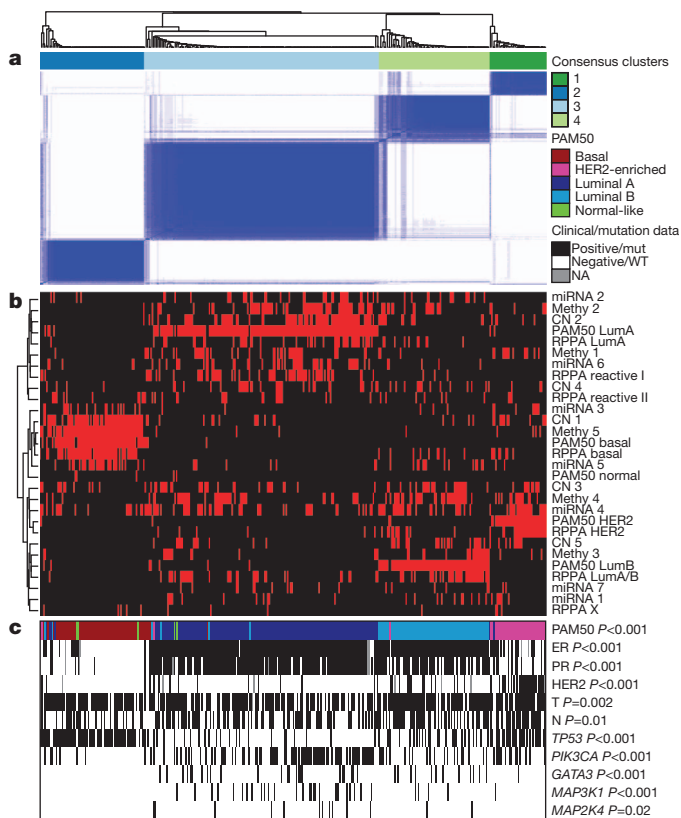


Figure 2 | Coordinated analysis of breast cancer subtypes defined from five different genomic/proteomic platforms. **a**, Consensus clustering analysis of the subtypes identifies four major groups (samples, $n = 348$). The blue and white heat map displays sample consensus. **b**, Heat-map display of the subtypes defined independently by miRNAs, DNA methylation, copy number (CN), PAM50 mRNA expression, and RPPA expression. The red bar indicates membership of a cluster type. **c**, Associations with molecular and clinical features. P values were calculated using a chi-squared test.

this classification in a C-of-C analysis instead of our five-class copy number subtypes, with either the PAM50 (Supplementary Fig. 15) or 12 unsupervised mRNA subtypes (Supplementary Fig. 16); each of these C-of-C classifications was highly correlated with PAM50 mRNA subtypes and with the other C-of-C analyses (Fig. 2). The transcriptional profiling and RPPA platforms demonstrated a high correlation with the consensus structure, indicating that the information content from copy number aberrations, miRNAs and methylation is captured at the level of gene expression and protein function.

Luminal/ER⁺ summary analysis

Luminal/ER⁺ breast cancers are the most heterogeneous in terms of gene expression (Supplementary Fig. 5), mutation spectrum (Fig. 1), copy number changes (Supplementary Fig. 9) and patient outcomes^{1,14}. One of the most dominant features is high mRNA and protein expression of the luminal expression signature (Supplementary Fig. 5), which contains *ESR1*, *GATA3*, *FOXA1*, *XBPI* and *MYB*; the luminal/ER⁺ cluster also contained the largest number of significantly mutated genes. Most notably, *GATA3* and *FOXA1* were mutated in a mutually exclusive fashion, whereas *ESR1* and *XBPI* were typically highly expressed but infrequently mutated. Mutations in *RUNX1* and its dimerization partner *C/EBP* may also have a role in aberrant ER signalling in luminal tumours, as *RUNX1* functions as an ER 'DNA tethering factor'³². PARADIGM³³ analysis comparing luminal versus basal-like cancers further emphasized the presence of a hyperactivated FOXA1-ER complex as a critical network hub differentiating these two tumour subtypes (Supplementary Fig. 17).

A confirmatory finding here was the high mutation frequency of *PIK3CA* in luminal/ER⁺ breast cancers^{34,35}. Through multiple

technology platforms, we examined possible relationships between *PIK3CA* mutation, *PTEN* loss, *INPP4B* loss and multiple gene and protein expression signatures of pathway activity. RPPA data demonstrated that pAKT, pS6 and p4EBP1, typical markers of phosphatidylinositol-3-OH kinase (PI(3)K) pathway activation, were not elevated in *PIK3CA*-mutated luminal A cancers; instead, they were highly expressed in basal-like and HER2E mRNA subtypes (the latter having frequent *PIK3CA* mutations) and correlated strongly with *INPP4B* and *PTEN* loss, and to a degree with *PIK3CA* amplification. Similarly, protein³⁶ and three mRNA signatures^{37–39} of PI(3)K pathway activation were enriched in basal-like over luminal A cancers (Fig. 3a). This apparent disconnect between the presence of *PIK3CA* mutations and biomarkers of pathway activation has been previously noted³⁶.

Another striking luminal/ER⁺ subtype finding was the frequent mutation of *MAP3K1* and *MAP2K4*, which represent two contiguous steps within the p38-JNK1 pathway^{24,40}. These mutations are predicted to be inactivating, with *MAP2K4* also a target of focal DNA loss in luminal tumours (Supplementary Fig. 9). To explore the possible interplay between *PIK3CA*, *MAP3K* and *MAP2K4* signalling, MEMO analysis⁴¹ was performed to identify mutually exclusive alterations targeting frequently altered genes likely to belong to the same pathway (Fig. 4). Across all breast cancers, MEMO identified a set of modules that highlight the differential activation events within the receptor tyrosine kinase (RTK)-PI(3)K pathway (Fig. 4a); mutations of *PIK3CA* were very common in luminal/ER⁺ cancers whereas *PTEN* loss was more common in basal-like tumours. Almost all *MAP3K1* and *MAP2K4* mutations were in luminal tumours, yet *MAP3K1* and *MAP2K4* appeared almost mutually exclusive relative to one another.

The TP53 pathway was differentially inactivated in luminal/ER⁺ breast cancers, with a low TP53 mutation frequency in luminal A (12%) and a higher frequency in luminal B (29%) cancers (Fig. 1). In addition to TP53 itself, a number of other pathway-inactivating events occurred including *ATM* loss and *MDM2* amplification (Figs 3b and 4b), both of which occurred more frequently within luminal B cancers. Gene expression analysis demonstrated that individual markers of functional TP53 (*GADD45A* and *CDKN1A*), and TP53 activity^{42,43} signatures, were highest in luminal A cancers (Fig. 3b). These data indicate that the TP53 pathway remains largely intact in luminal A cancers but is often inactivated in the more aggressive luminal B cancers⁴⁴. Other PARADIGM-based pathway differences driving luminal B versus luminal A included hyperactivation of transcriptional activity associated with *MYC* and *FOXM1* proliferation.

The critical retinoblastoma/RB1 pathway also showed mRNA-subtype-specific alterations (Fig. 3c). RB1 itself, by mRNA and protein expression, was detectable in most luminal cancers, with highest levels within luminal A. A common oncogenic event was cyclin D1 amplification and high expression, which preferentially occurred within luminal tumours, and more specifically within luminal B. In contrast, the presumed tumour suppressor *CDKN2C* (also called *p18*) was at its lowest levels in luminal A cancers, consistent with observations in mouse models⁴⁵. Finally, RB1 activity signatures were also high in luminal cancers^{46–48}. Luminal A tumours, which have the best prognosis, are the most likely to retain activity of the major tumour suppressors RB1 and TP53.

These genomic characterizations also provided clues for druggable targets. We compiled a drug target table in which we defined a target as a gene/protein for which there is an approved or investigational drug in human clinical trials targeting the molecule or canonical pathway (Supplementary Table 6). In luminal/ER⁺ cancers, the high frequency of *PIK3CA* mutations suggests that inhibitors of this activated kinase or its signalling pathway may be beneficial. Other potential significantly mutated gene drug candidates include AKT1 inhibitors (11 out of 12 *AKT1* variants were luminal) and PARP inhibitors for *BRCA1/BRCA2* mutations. Although still unapproved as biomarkers, many potential copy-number-based drug targets

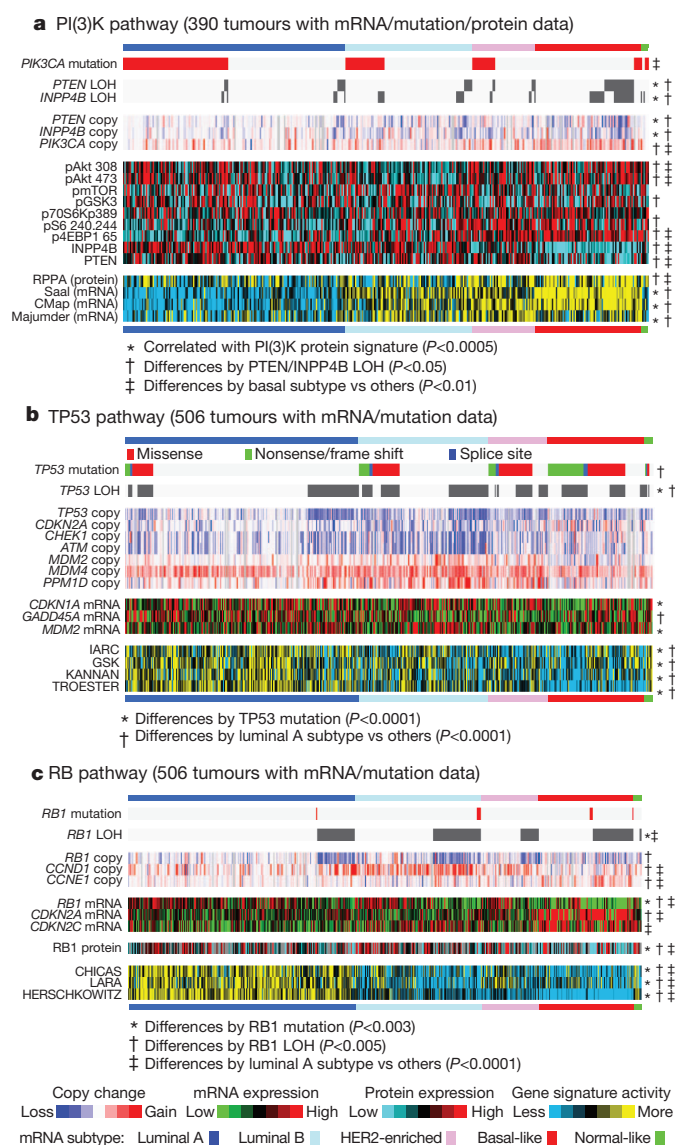


Figure 3 | Integrated analysis of the PI(3)K, TP53 and RB1 pathways. Breast cancer subtypes differ by genetic and genomic targeting events, with corresponding effects on pathway activity. **a–c**, For PI(3)K (**a**), TP53 (**b**) and RB1 (**c**) pathways, key genes were selected using prior biological knowledge. Multiple mRNA expression signatures for a given pathway were defined (details in Supplementary Methods; PI(3)K:Saal, PTEN loss in human breast tumours; CMap, PI(3)K/mTOR inhibitor treatment *in vitro*; Majumder, Akt overexpression in mouse model; TP53: IARC, expert-curated p53 targets; GSK, TP53 mutant versus wild-type cell lines; KANNAN, TP53 overexpression *in vitro*; TROESTER, TP53 knockdown *in vitro*; RB: CHICAS, RB1 mouse knockout versus wild type; LARA, RB1 knockdown *in vitro*; HERSCHKOWITZ, RB1 loss of heterozygosity (LOH) in human breast tumours) and applied to the gene expression data, in order to score each tumour for relative signature activity (yellow, more active). The PI(3)K panel includes a protein-based (RPPA) proteomic signature. Tumours were ordered first by mRNA subtype, although specific ordering differs between the panels. P values were calculated by a Pearson's correlation or a Chi-squared test.

were identified including amplifications of fibroblast growth factor receptors (FGFRs) and *IGFR1*, as well as cyclin D1, *CDK4* and *CDK6*. A summary of the general findings in luminal tumours and the other subtypes is presented in Table 1.

HER2-based classifications and summary analysis

DNA amplification of *HER2* was readily evident in this study (Supplementary Fig. 9) together with overexpression of multiple

HER2-amplicon-associated genes that in part define the HER2E mRNA subtype (Supplementary Fig. 5). However, not all clinically HER2⁺ tumours are of the HER2E mRNA subtype, and not all tumours in the HER2E mRNA subtype are clinically HER2⁺. Integrated analysis of the RPPA and mRNA data clearly identified a HER2⁺ group (Supplementary Fig. 12). When the HER2⁺ protein and HER2E mRNA subtypes overlapped, a strong signal of EGFR, pEGFR, HER2 and pHER2 was observed. However, only ~50% of clinically HER2⁺ tumours fall into this HER2E-mRNA-subtype/HER2-protein group, the rest of the clinically HER2⁺ tumours were observed predominantly in the luminal mRNA subtypes.

These data indicate that there exist at least two types of clinically defined HER2⁺ tumours. To identify differences between these groups, a supervised gene expression analysis comparing 36 HER2E-mRNA-subtype/HER2⁺ versus 31 luminal-mRNA-subtype/HER2⁺ tumours was performed and identified 302 differentially expressed genes (q -value = 0%) (Supplementary Fig. 18 and Supplementary Table 7). These genes largely track with ER status but also indicated that HER2E-mRNA-subtype/HER2⁺ tumours showed significantly higher expression of a number of RTKs including *FGFR4*, *EGFR*, *HER2* itself, as well as genes within the HER2 amplicon (including *GRB7*). Conversely, the luminal-mRNA-subtype/HER2⁺ tumours showed higher expression of the luminal cluster of genes including *GATA3*, *BCL2* and *ESR1*. Further support for two types of clinically defined HER2⁺ disease was evident in the somatic mutation data supervised by either mRNA subtype or ER status; TP53 mutations were significantly enriched in HER2E or ER-negative tumours whereas *GATA3* mutations were only observed in luminal subtypes or ER⁺ tumours.

Analysis of the RPPA data according to mRNA subtype identified 36 differentially expressed proteins (q -value < 5%) (Supplementary Fig. 18G and Supplementary Table 8). The EGFR/pEGFR/HER2/pHER2 signal was again observed and present within the HER2E-mRNA-subtype/HER2⁺ tumours, as was high pSRC and pS6; conversely, many protein markers of luminal cancers again distinguished the luminal-mRNA-subtype/HER2⁺ tumours. Given the importance of clinical HER2 status, a more focused analysis was performed based on the RPPA-defined protein expression of HER2 (Supplementary Fig. 19)—the results strongly recapitulated findings from the RPPA and mRNA subtypes including a high correlation between HER2 clinical status, HER2 protein by RPPA, pHER2, EGFR and pEGFR. These multiple signatures, namely HER2E mRNA subtype, HER2 amplicon genes by mRNA expression, and RPPA EGFR/pEGFR/HER2/pHER2 signature, ultimately identify at least two groups/subtypes within clinically HER2⁺ tumours (Table 1). These signatures represent breast cancer biomarker(s) that could potentially predict response to anti-HER2 targeted therapies.

Many therapeutic advances have been made for clinically HER2⁺ disease. This study has identified additional somatic mutations that represent potential therapeutic targets within this group, including a high frequency of *PIK3CA* mutations (39%), a lower frequency of *PTEN* and *PIK3R1* mutations (Supplementary Table 6), and genomic losses of *PTEN* and *INPP4B*. Other possible druggable mutations included variants within HER family members including two somatic mutations in *HER2*, two within *EGFR*, and five within *HER3*. Pertuzumab, in combination with trastuzumab, targets the HER2–HER3 heterodimer⁴⁹; however, these data suggest that targeting EGFR with HER2 could also be beneficial. Finally, the HER2E mRNA subtype typically showed high aneuploidy, the highest somatic mutation rate (Table 1), and DNA amplification of other potential therapeutic targets including FGFRs, *EGFR*, *CDK4* and cyclin D1.

Basal-like summary analysis

The basal-like subtype was discovered more than a decade ago by first-generation cDNA microarrays¹³. These tumours are often referred to as triple-negative breast cancers (TNBCs) because most basal-like tumours are typically negative for ER, PR and HER2.

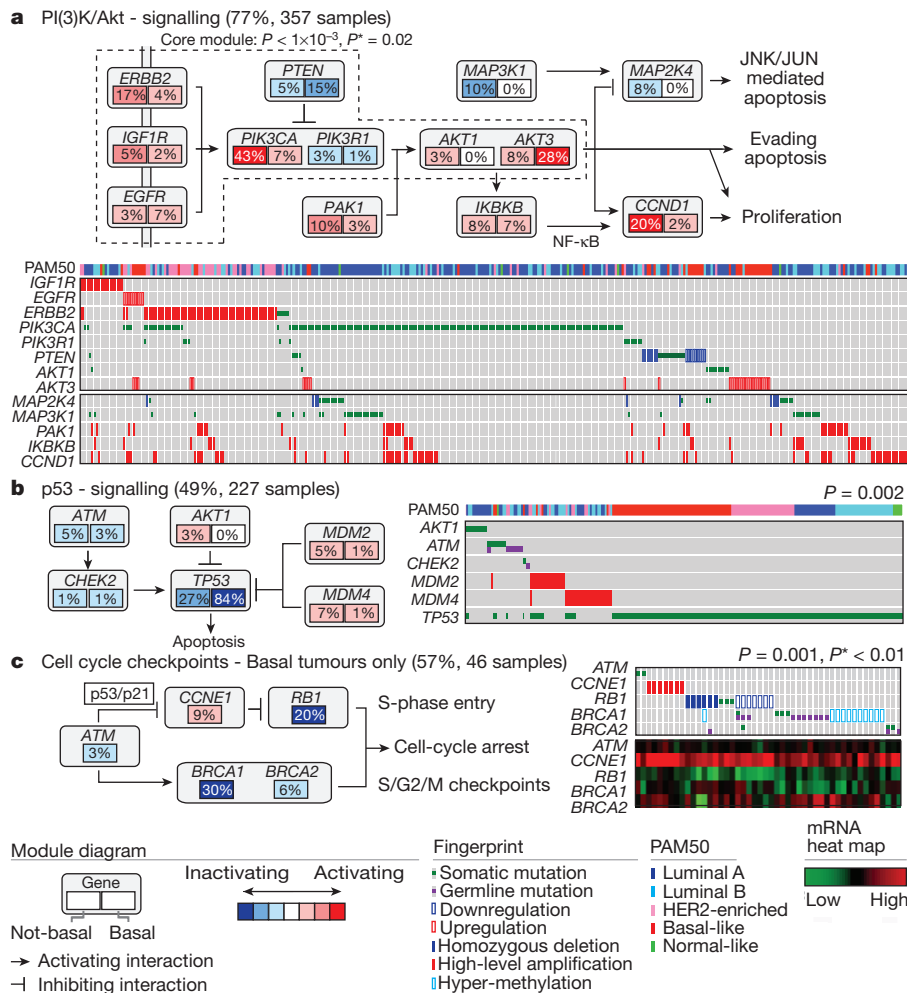


Figure 4 | Mutual exclusivity modules in cancer (MEMo) analysis. Mutual exclusivity modules are represented by their gene components and connected to reflect their activity in distinct pathways. For each gene, the frequency of alteration in basal-like (right box) and non-basal (left box) is reported. Next to each module is a fingerprint indicating what specific alteration is observed for each gene (row) in each sample (column). **a**, MEMo identified several overlapping modules that recapitulate the RTK-PI(3)K and p38-JNK1

signalling pathways and whose core was the top-scoring module. **b**, MEMo identified alterations to TP53 signalling as occurring within a statistically significant mutually exclusive trend. **c**, A basal-like only MEMo analysis identified one module that included ATM mutations, defects at *BRCA1* and *BRCA2*, and deregulation of the *RB1* pathway. A gene expression heat map is below the fingerprint to show expression levels.

However, ~75% of TNBCs are basal-like with the other 25% comprised of all other mRNA subtypes⁶. In this data set, there was a high degree of overlap between these two distinctions with 76 TNBCs, 81 basal-like, and 65 that were both TNBCs and basal-like. Given the known heterogeneity of TNBCs, and that the basal-like subtype proved to be distinct on every platform, we chose to use the basal-like distinction for comparative analyses.

Basal-like tumours showed a high frequency of *TP53* mutations (80%)⁹, which when combined with inferred *TP53* pathway activity suggests that loss of *TP53* function occurs within most, if not all, basal-like cancers (Fig. 3b). In addition to loss of *TP53*, a MEMo analysis reconfirmed that loss of *RB1* and *BRCA1* are basal-like features (Fig. 4c)^{47,50}. *PIK3CA* was the next most commonly mutated gene (~9%); however, inferred PI(3)K pathway activity, whether from gene^{37–39}, protein³⁶, or high PI(3)K/AKT pathway activities, was highest in basal-like cancers (Fig. 3a). Alternative means of activating the PI(3)K pathway in basal-like cancers probably includes loss of *PTEN* and *INPP4B* and/or amplification of *PIK3CA*. A recent paper¹² performed exome sequencing of 102 TNBCs. Five of the top six most frequent TNBC mutations in ref. 12 were also observed at a similar frequency in our TNBC subset (*Myo3A* not present here); of those five, three passed our test as a significantly mutated gene in TNBCs (Supplementary Table 2).

Expression features of basal-like tumours include a characteristic signature containing keratins 5, 6 and 17 and high expression of genes associated with cell proliferation (Supplementary Fig. 5). A PARADIGM³³ analysis of basal-like versus luminal tumours emphasized the importance of hyperactivated FOXM1 as a transcriptional driver of this enhanced proliferation signature (Supplementary Fig. 17). PARADIGM also identified hyperactivated MYC and HIF1- α /ARNT network hubs as key regulatory features of basal-like cancers. Even though chromosome 8q24 is amplified across all subtypes (Supplementary Fig. 9), high MYC activation seems to be a basal-like characteristic⁵¹.

Given the striking contrasts between basal-like and luminal/HER2E subtypes, we performed a MEMo analysis on basal-like tumours alone. The top-scoring module included *ATM* mutations, *BRCA1* and *BRCA2* inactivation, *RB1* loss and cyclin E1 amplification (Fig. 4c). Notably, these same modules were identified previously for serous ovarian cancers⁴¹. Furthermore, the basal-like (and TNBC) mutation spectrum was reminiscent of the spectrum seen in serous ovarian cancers⁵² with only one gene (that is, *TP53*) at >10% mutation frequency. To explore possible similarities between serous ovarian and the breast basal-like cancers, we performed a number of analyses comparing ovarian versus breast luminal, ovarian versus breast basal-like, and breast basal-like versus breast luminal cancers

Table 1 | Highlights of genomic, clinical and proteomic features of subtypes

Subtype	Luminal A	Luminal B	Basal-like	HER2E
ER ⁺ /HER2 ⁻ (%)	87	82	10	20
HER2 ⁺ (%)	7	15	2	68
TNBCs (%)	2	1	80	9
TP53 pathway	<i>TP53</i> mut (12%); gain of <i>MDM2</i> (14%)	<i>TP53</i> mut (32%); gain of <i>MDM2</i> (31%)	<i>TP53</i> mut (84%); gain of <i>MDM2</i> (14%)	<i>TP53</i> mut (75%); gain of <i>MDM2</i> (30%)
PIK3CA/PTEN pathway	<i>PIK3CA</i> mut (49%); <i>PTEN</i> mut/loss (13%); <i>INPP4B</i> loss (9%)	<i>PIK3CA</i> mut (32%); <i>PTEN</i> mut/loss (24%); <i>INPP4B</i> loss (16%)	<i>PIK3CA</i> mut (7%); <i>PTEN</i> mut/loss (35%); <i>INPP4B</i> loss (30%)	<i>PIK3CA</i> mut (42%); <i>PTEN</i> mut/loss (19%); <i>INPP4B</i> loss (30%)
RB1 pathway	Cyclin D1 amp (29%); <i>CDK4</i> gain (14%); low expression of <i>CDKN2C</i> ; high expression of <i>RB1</i>	Cyclin D1 amp (58%); <i>CDK4</i> gain (25%)	<i>RB1</i> mut/loss (20%); cyclin E1 amp (9%); high expression of <i>CDKN2A</i> ; low expression of <i>RB1</i>	Cyclin D1 amp (38%); <i>CDK4</i> gain (24%)
mRNA expression	High ER cluster; low proliferation	Lower ER cluster; high proliferation	Basal signature; high proliferation	HER2 amplicon signature; high proliferation
Copy number	Most diploid; many with quiet genomes; 1q, 8q, 8p11 gain; 8p, 16q loss; 11q13.3 amp (24%)	Most aneuploid; many with focal amp; 1q, 8q, 8p11 gain; 8p, 16q loss; 11q13.3 amp (51%); 8p11.23 amp (28%)	Most aneuploid; high genomic instability; 1q, 10p gain; 8p, 5q loss; <i>MYC</i> focal gain (40%)	Most aneuploid; high genomic instability; 1q, 8q gain; 8p loss; 17q12 focal <i>ERBB2</i> amp (71%)
DNA mutations	<i>PIK3CA</i> (49%); <i>TP53</i> (12%); <i>GATA3</i> (14%); <i>MAP3K1</i> (14%)	<i>TP53</i> (32%); <i>PIK3CA</i> (32%); <i>MAP3K1</i> (5%)	<i>TP53</i> (84%); <i>PIK3CA</i> (7%)	<i>TP53</i> (75%); <i>PIK3CA</i> (42%); <i>PIK3R1</i> (8%)
DNA methylation	–	Hypermethylated phenotype for subset	Hypomethylated	–
Protein expression	High oestrogen signalling; high MYB; RPPA reactive subtypes	Less oestrogen signalling; high FOXM1 and MYC; RPPA reactive subtypes	High expression of DNA repair proteins, <i>PTEN</i> and <i>INPP4B</i> loss signature (pAKT)	High protein and phospho-protein expression of EGFR and HER2

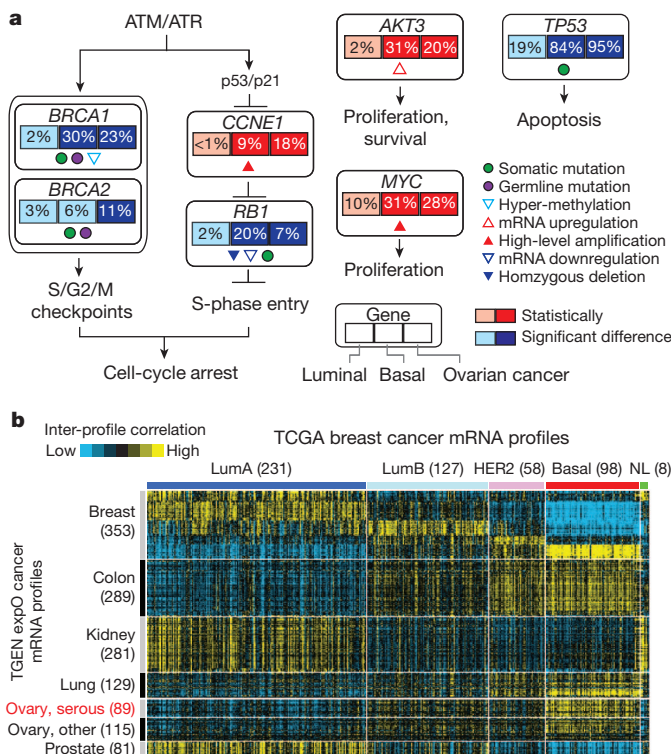
Percentages are based on 466 tumour overlap list. Amp, amplification; mut, mutation.

(Fig. 5). Comparing copy number landscapes, we observed several common features between ovarian and basal-like tumours including widespread genomic instability and common gains of 1q, 3q, 8q and 12p, and loss of 4q, 5q and 8p (Supplementary Fig. 20A). Using a more global copy number comparison, we examined the overall fraction of the genome altered and the overall copy number correlation of ovarian cancers versus each breast cancer mRNA subtype

(Supplementary Fig. 20A, B); in both cases, basal-like tumours were the most similar to the serous ovarian carcinomas.

We systematically looked for other common features between serous ovarian and basal-like tumours when each was compared to luminal. We identified: (1) *BRCA1* inactivation; (2) *RB1* loss and cyclin E1 amplification; (3) high expression of *AKT3*; (4) *MYC* amplification and high expression; and (5) a high frequency of *TP53* mutations (Fig. 5a). An additional supervised analysis of a large, external multitumour type transcriptomic data set (Gene Expression Omnibus accession GSE2109) was performed where each TCGA (The Cancer Genome Atlas) breast tumour expression profile was compared via a correlation analysis to that of each tumour in the multitumour set. Basal-like breast cancers clearly showed high mRNA expression correlations with serous ovarian cancers, as well as with lung squamous carcinomas (Fig. 5b). A PARADIGM analysis that calculates whether a gene or pathway feature is both differentially activated in basal-like versus luminal cancers and has higher overall activity across the TCGA ovarian samples was performed; this identified comparably high pathway activity of the HIF1- α /ARNT, MYC and FOXM1 regulatory hubs in both ovarian and basal-like cancers (Supplementary Fig. 20C). The common findings of *TP53*, *RB1* and *BRCA1* loss, with *MYC* amplification, strongly suggest that these are shared driving events for basal-like and serous ovarian carcinogenesis. This suggests that common therapeutic approaches should be considered, which is supported by the activity of platinum analogues and taxanes in breast basal-like and serous ovarian cancers.

Given that most basal-like cancers are TNBCs, finding new drug targets for this group is critical. Unfortunately, the somatic mutation repertoire for basal-like breast cancers has not provided a common target aside from *BRCA1* and *BRCA2*. Here we note that ~20% of basal-like tumours had a germline ($n = 12$) and/or somatic ($n = 8$) *BRCA1* or *BRCA2* variant, which suggests that one in five basal-like patients might benefit from PARP inhibitors and/or platinum compounds^{53,54}. The copy number landscape of basal-like cancers showed multiple amplifications and deletions, some of which may provide therapeutic targets (Supplementary Table 6). Potential targets include losses of *PTEN* and *INPP4B*, both of which have been shown to sensitize cell lines to PI(3)K pathway inhibitors^{55,56}. Interestingly, many of the components of the PI(3)K and RAS–RAF–MEK pathway were amplified (but not typically mutated) in basal-like cancers including *PIK3CA* (49%), *KRAS* (32%), *BRAF* (30%) and *EGFR* (23%). Other RTKs that are plausible drug targets and amplified in

**Figure 5 | Comparison of breast and serous ovarian carcinomas.**

a, Significantly enriched genomic alterations identified by comparing basal-like or serous ovarian tumours to luminal cancers. **b**, Inter-sample correlations (yellow, positive) between gene transcription profiles of breast tumours (columns; TCGA data, arranged by subtype) and profiles of cancers from various tissues of origin (rows; external 'TGEN expO' data set, GSE2109) including ovarian cancers.

some basal-like cancers include *FGFR1*, *FGFR2*, *IGFR1*, *KIT*, *MET* and *PDGFRA*. Finally, the PARADIGM identification of high HIF1- α /ARNT pathway activity suggests that these malignancies might be susceptible to angiogenesis inhibitors and/or bioreductive drugs that become activated under hypoxic conditions.

Concluding remarks

The integrated molecular analyses of breast carcinomas that we report here significantly extends our knowledge base to produce a comprehensive catalogue of likely genomic drivers of the most common breast cancer subtypes (Table 1). Our novel observation that diverse genetic and epigenetic alterations converge phenotypically into four main breast cancer classes is not only consistent with convergent evolution of gene circuits, as seen across multiple organisms, but also with models of breast cancer clonal expansion and *in vivo* cell selection proposed to explain the phenotypic heterogeneity observed within defined breast cancer subtypes.

METHODS SUMMARY

Specimens were obtained from patients with appropriate consent from institutional review boards. Using a co-isolation protocol, DNA and RNA were purified. In total, 800 patients were assayed on at least one platform. Different numbers of patients were used for each platform using the largest number of patients available at the time of data freeze; 466 samples (463 patients) were in common across 5 out of 6 platforms (excluding RPPA) and 348 patients were in common on 6 out of 6 platforms. Technology platforms used include: (1) gene expression DNA microarrays⁵²; (2) DNA methylation arrays; (3) miRNA sequencing; (4) Affymetrix SNP arrays; (5) exome sequencing; and (6) reverse phase protein arrays. Each platform, except for the exome sequencing, was used in a *de novo* subtype discovery analysis (Supplementary Methods) and then included in a single analysis to define an overall subtype architecture. Additional integrated across-platform computational analyses were performed including PARADIGM³³ and MEMo⁴¹.

Received 22 March; accepted 11 July 2012.

Published online 23 September; corrected online 3 October 2012.

- Paik, S. *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).
- van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
- Slamon, D. J. *et al.* Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177–182 (1987).
- Chin, K. *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10**, 529–541 (2006).
- Bergamaschi, A. *et al.* Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosomes. Cancer* **45**, 1033–1040 (2006).
- Perou, C. M. Molecular stratification of triple-negative breast cancers. *Oncologist* **16** (suppl. 1), 61–70 (2011).
- Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl Acad. Sci. USA* **100**, 8418–8423 (2003).
- Foulkes, W. D. *et al.* Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *J. Natl Cancer Inst.* **95**, 1482–1485 (2003).
- Carey, L. A. *et al.* Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *J. Am. Med. Assoc.* **295**, 2492–2502 (2006).
- Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**, 999–1005 (2010).
- Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
- Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**, 395–399 (2012).
- Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
- González-Pérez, A. & López-Bigas, N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am. J. Hum. Genet.* **88**, 440–449 (2011).
- Dees, N. D. *et al.* MuSiC: Identifying mutational significance in cancer genomes. *Genome Res.* **22**, 1589–1598 (2012).
- Bamshad, M. *et al.* Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nature Genet.* **16**, 311–315 (1997).
- Li, Q. Y. *et al.* Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nature Genet.* **15**, 21–29 (1997).
- The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
- Cheung, L. W. *et al.* High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. *Cancer Discov.* **1**, 170–185 (2011).
- Malcovati, L. *et al.* Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood* **118**, 6239–6246 (2011).
- Wang, L. *et al.* SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365**, 2497–2506 (2011).
- Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075 (2008).
- Johnson, G. L. & Lapadat, R. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* **298**, 1911–1912 (2002).
- Usary, J. *et al.* Mutation of GATA3 in human breast tumors. *Oncogene* **23**, 7669–7678 (2004).
- Walsh, T. *et al.* Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc. Natl Acad. Sci. USA* **107**, 12629–12633 (2010).
- Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* **12**, R68 (2010).
- Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152–D157 (2011).
- Weigman, V. J. *et al.* Basal-like breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival. *Breast Cancer Res. Treat.* **133**, 865–880 (2011).
- Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
- Hennessy, B. T. *et al.* A technical assessment of the utility of reverse phase protein arrays for the study of the functional proteome in non-microdissected human breast cancers. *Clin. Proteomics* **6**, 129–151 (2010).
- Daub, H. *et al.* Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol. Cell* **31**, 438–448 (2008).
- Vaske, C. J. *et al.* Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26**, 237–245 (2010).
- Campbell, I. G. *et al.* Mutation of the PIK3CA gene in ovarian and breast cancer. *Cancer Res.* **64**, 7678–7681 (2004).
- Bachman, K. E. *et al.* The PIK3CA gene is mutated with high frequency in human breast cancers. *Cancer Biol. Ther.* **3**, 772–775 (2004).
- Stemke-Hale, K. *et al.* An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. *Cancer Res.* **68**, 6084–6091 (2008).
- Creighton, C. J. *et al.* Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER⁺ breast cancer. *Breast Cancer Res.* **12**, R40 (2010).
- Majumder, P. K. *et al.* mTOR inhibition reverses Akt-dependent prostate intraepithelial neoplasia through regulation of apoptotic and HIF-1-dependent pathways. *Nature Med.* **10**, 594–601 (2004).
- Saal, L. H. *et al.* Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair. *Nature Genet.* **40**, 102–107 (2008).
- Wagner, E. F. & Nebreda, A. R. Signal integration by JNK and p38 MAPK pathways in cancer development. *Nature Rev. Cancer* **9**, 537–549 (2009).
- Ciriello, G., Cerami, E., Sander, C. & Schultz, N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res.* **22**, 398–406 (2012).
- Kannan, K. *et al.* DNA microarrays identification of primary and secondary target genes regulated by p53. *Oncogene* **20**, 2225–2234 (2001).
- Troester, M. A. *et al.* Gene expression patterns associated with p53 status in breast cancer. *BMC Cancer* **6**, 276 (2006).
- Deisenroth, C., Thorner, A. R., Enomoto, T., Perou, C. M. & Zhang, Y. Mitochondrial Hep27 is a c-Myb target gene that inhibits Mdm2 and stabilizes p53. *Mol. Cell. Biol.* **30**, 3981–3993 (2010).
- Pei, X. H. *et al.* CDK inhibitor p18^{INK4c} is a downstream target of GATA3 and restrains mammary luminal progenitor cell proliferation and tumorigenesis. *Cancer Cell* **15**, 389–401 (2009).
- Chicas, A. *et al.* Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17**, 376–387 (2010).
- Herschkowitz, J. I., He, X., Fan, C. & Perou, C. M. The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal B breast carcinomas. *Breast Cancer Res.* **10**, R75 (2008).
- Lara, M. F. *et al.* Gene profiling approaches help to define the specific functions of retinoblastoma family in epidermis. *Mol. Carcinog.* **47**, 209–221 (2008).
- Baselga, J. *et al.* Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N. Engl. J. Med.* **366**, 109–119 (2012).
- Jiang, Z. *et al.* Rb deletion in mouse mammary progenitors induces luminal-B or basal-like/EMT tumor subtypes depending on p53 status. *J. Clin. Invest.* **120**, 3296–3309 (2010).
- Chandriani, S. *et al.* A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response. *PLoS ONE* **4**, e6693 (2009).
- The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
- Audeh, M. W. *et al.* Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet* **376**, 245–251 (2010).
- Fong, P. C. *et al.* Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.* **361**, 123–134 (2009).

55. Fedele, C. G. *et al.* Inositol polyphosphate 4-phosphatase II regulates PI3K/Akt signaling and is lost in human basal-like breast cancers. *Proc. Natl Acad. Sci. USA* **107**, 22231–22236 (2010).
56. Gewinner, C. *et al.* Evidence that inositol polyphosphate 4-phosphatase type II is a tumor suppressor that inhibits PI3K signaling. *Cancer Cell* **16**, 115–125 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Sheth and S. Lucas for administrative coordination of TCGA activities, and C. Gunter for critical reading of the manuscript. This work was supported by the following grants from the USA National Institutes of Health: U24CA143883, U24CA143858, U24CA143840, U24CA143799, U24CA143835, U24CA143845, U24CA143882, U24CA143867, U24CA143866, U24CA143848, U24CA144025, U54HG003079, P50CA116201 and P50CA58223. Additional support was provided by the Susan G. Komen for the Cure, the US Department of Defense through the Henry M. Jackson Foundation for the Advancement of Military Medicine, and the Breast Cancer Research Foundation. The views expressed in this paper are those of the authors and do not reflect the official policy of the Department of Defense, or US Government.

Author Contributions TCGA research network contributed collectively to this study. Biospecimens were provided by tissue source sites and processed by a Biospecimens Core Resource. Data generation and analyses were performed by genome sequencing centres, cancer genome characterization centres, and genome data analysis centres. RPPA analysis was performed at the MD Anderson Cancer Center in association with the genome data analysis centre. All data were released through the Data Coordinating Center. Project activities were coordinated by NCI and NHGRI project teams. We also acknowledge the following TCGA investigators of the Breast Analysis Working Group who contributed substantially to the analysis and writing of this manuscript: Project leaders, C.M.P., M.J.E.; manuscript coordinator, C.M.P., K.A.H.; data coordinator, K.A.H.; analysis coordinator, C.M.P., K.A.H.; DNA sequence analysis, D.C.K., A.D.; mRNA microarray analysis, K.A.H., C.F.; mRNA sequence analysis, A.G.R., A.C.; DNA methylation analysis, S. Malik, S. Mahurkar, P.W.L.; copy number analysis, A.D.C., M.M.; protein analysis, W.L., R.G.W.V., G.B.M.; pathway/integrated analysis, C.J.C., C.Y., J.M.S., C.C.B., G.C., C.S., S.R., I.S.; biospecimen core resource, T.L., J.B., J.M.G.; pathology and clinical expertise, T.A.K., H.H., R.J.M., J.N.I., T.S., F.W.

Author Information All of the primary sequence files are deposited in CGHub (<https://cghub.ucsc.edu/>); all other data including mutation annotation file are deposited at the Data Coordinating Center (<http://cancergenome.nih.gov/>). Sample lists, data matrices and supporting data can be found at http://tcga-data.nci.nih.gov/docs/publications/brca_2012/. The data can be explored via the ISB Regulome Explorer (<http://explorer.cancerregulome.org/>) and the cBio Cancer Genomics Portal (<http://cbioportal.org>). Data descriptions can be found at <https://wiki.nci.nih.gov/display/TCGA/TCGA+Data+Primer> and in Supplementary Methods. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and the online version of this paper is freely available to all readers. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.M.P. (cperou@med.unc.edu).

The Cancer Genome Atlas Network

Genome sequencing centres: Washington University in St Louis Daniel C. Koboldt¹, Robert S. Fulton¹, Michael D. McLellan¹, Heather Schmidt¹, Joelle Kalicki-Verizer¹, Joshua F. McMichael¹, Lucinda L. Fulton¹, David J. Dooling¹, Li Ding^{1,2}, Elaine R. Mardis^{1,2,3}, Richard K. Wilson^{1,2,3}

Genome characterization centres: BC Cancer Agency Adrian Ally⁴, Miruna Balasundaram⁴, Yaron S. N. Butterfield⁴, Rebecca Carlsen⁴, Candace Carter⁴, Andy Chu⁴, Eric Chuah⁴, Hye-Jung E. Chun⁴, Robin J. N. Coope⁴, Noreen Dhalla⁴, Ranabir Guin⁴, Carrie Hirst⁴, Martin Hirst⁴, Robert A. Holt⁴, Darlene Lee⁴, Haiyan I. Li⁴, Michael Mayo⁴, Richard A. Moore⁴, Andrew J. Mungall⁴, Erin Pleasance⁴, A. Gordon Robertson⁴, Jacqueline E. Schein⁴, Arash Shafiei⁴, Payal Sipahimalani⁴, Jared R. Slobodan⁴, Dominik Stoll⁴, Angela Tam⁴, Nina Thiessen⁴, Richard J. Varhol⁴, Natasja Wye⁴, Thomas Zeng⁴, Yongjun Zhao⁴, Inanc Birol⁴, Steven J. M. Jones⁴, Marco A. Marra⁴; **Broad Institute** Andrew D. Cherniack⁵, Gordon Saksena⁵, Robert C. Onofrio⁵, Nam H. Pho⁵, Scott L. Carter⁵, Steven E. Schumacher^{5,6}, Barbara Tabak^{5,6}, Bryan Hernandez⁵, Jeff Gentry⁵, Huy Nguyen⁵, Andrew Crenshaw⁵, Kristin Ardlie⁵, Rameen Beroukhi^{5,7,8}, Wendy Winkler⁵, Gad Getz⁵, Stacey B. Gabriel⁵, Matthew Meyerson^{5,9,10}; **Brigham & Women's Hospital & Harvard Medical School** Lynda Chin^{9,11}, Peter J. Park¹², Raju Kucherlapati¹³; **University of North Carolina, Chapel Hill** Katherine A. Hoadley^{14,15}, J. Todd Auman^{16,17}, Cheng Fan¹⁵, Yidi J. Turman¹⁵, Yan Shi¹⁵, Ling Li¹⁵, Michael D. Topal^{15,18}, Xiaping He^{14,15}, Hann-Hsiang Chao^{14,15}, Aleix Prat^{14,15}, Grace O. Silva^{14,15}, Michael D. Iglesia^{14,15}, Wei Zhao^{14,15}, Jerry Usary¹⁵, Jonathan S. Berg^{14,15}, Michael Adams¹⁴, Jessica Booker¹⁸, Junyuan Wu¹⁵, Anisha Gulabani¹⁵, Tom Bodenheimer¹⁵, Alan P. Hoyle¹⁵, Janae V. Simons¹⁵, Matthew G. Soloway¹⁵, Lisle E. Mose¹⁵, Stuart R. Jefferys¹⁵, Saianand Balu¹⁵, Joel S. Parker¹⁵, D. Neil Hayes^{15,19}, Charles M. Perou^{14,15,18}; **University of Southern California/Johns Hopkins** Simeen Malik²⁰, Swapna Mahurkar²⁰, Hui Shen²⁰, Daniel J. Weisenberger²⁰, Timothy Triche Jr²⁰, Phillip H. Lai²⁰, Moiz S. Bootwalla²⁰, Dennis T. Maglinte²⁰, Benjamin P. Berman²⁰, David J. Van Den Berg²⁰, Stephen B. Baylin²¹, Peter W. Laird²⁰

Genome data analysis: Baylor College of Medicine Chad J. Creighton^{22,23}, Lawrence A. Donehower^{22,23,24,25}; **Broad Institute** Gad Getz²⁶, Michael Noble²⁶, Doug Voet²⁶, Gordon Saksena²⁶, Nils Gehlenborg^{12,26}, Daniel DiCara²⁶, Junhua Zhang²⁷, Hailei Zhang²⁶, Chang-Jiun Wu²⁸, Spring Yingchun Liu²⁸, Michael S. Lawrence²⁶, Lihua Zou²⁶, Andrey Sivachenko²⁶, Pei Lin²⁶, Petar Stojanov²⁶, Rui Jing²⁶, Juok Cho²⁶, Raktim Sinha²⁶, Richard W. Park²⁶, Marc-Danie Nazaire²⁶, Jim Robinson²⁶, Helga Thorvaldsdottir²⁶, Jill Mesirov²⁶, Peter J. Park^{12,29,30}, Lynda Chin^{26,27}; **Institute for Systems Biology** Sheila Reynolds³¹, Richard B. Kreiberg³¹, Brady Bernard³¹, Ryan Bressler³¹, Timo Erkkila³², Jake Lin³¹, Vestinn Thorsson³¹, Wei Zhang³³, Ilya Shmulevich³¹; **Memorial Sloan-Kettering Cancer Center** Giovanni Ciriello³⁴, Nils Weinhold³⁴, Nikolaus Schultz³⁴, Jianjiong Gao³⁴, Ethan Cerami³⁴, Benjamin Gross³⁴, Anders Jacobsen³⁴, Rileen Sinha³⁴, B. Arman Aksoy³⁴, Yevgeniy Antipin³⁴, Boris Reva³⁴, Ronglai Shen³⁵, Barry S. Taylor³⁴, Marc Ladanyi³⁶, Chris Sander³⁴; **Oregon Health & Science University** Pavana Anur³⁷, Paul T. Spellman³⁷; **The University of Texas MD Anderson Cancer Center** Yiling Lu^{38,39}, Wenbin Liu⁴⁰, Roel R. G. Verhaak⁴⁰, Gordon B. Mills^{38,39}, Rehan Akbani⁴⁰, Nianxiang Zhang⁴⁰, Bradley M. Broom⁴⁰, Tod D. Casasent⁴⁰, Chris Wakefield⁴⁰, Anna K. Unruh⁴⁰, Keith Baggerly⁴⁰, Kevin Coombes⁴⁰, John N. Weinstein⁴⁰; **University of California, Santa Cruz/Buck Institute** David Haussler^{41,42}, Christopher C. Benz⁴³, Joshua M. Stuart⁴¹, Stephen C. Benz⁴¹, Jingchun Zhu⁴¹, Christopher C. Szeto⁴¹, Gary K. Scott⁴³, Christina Yu⁴³, Evan O. Paull⁴¹, Daniel Carlin⁴¹, Christopher Wong⁴¹, Artem Sokolov⁴¹, Janita Thusberg⁴³, Sean Mooney⁴³, Sam Ng⁴¹, Theodore C. Goldstein⁴¹, Kyle Ellrott⁴¹, Mia Griffo⁴¹, Christopher Wilks⁴¹, Singer Ma⁴¹, Brian Craft⁴¹; **NCI Chunhua Yan**⁴⁴, Ying Hu⁴⁴, Daoud Meerzaman⁴⁴

Biospecimen core resource: Nationwide Children's Hospital Biospecimen Core Resource Julie M. Gastier-Foster^{45,46,47}, Jay Bowen⁴⁷, Nilsa C. Ramirez^{45,47}, Aaron D. Black⁴⁷, Robert E. Pyatt^{45,47}, Peter White^{46,47}, Erik J. Zmuda⁴⁷, Jessica Frick⁴⁷, Tara M. Lichtenberg⁴⁷, Robin Brokens⁴⁷, Myra M. George⁴⁷, Mark A. Gerken⁴⁷, Hollie A. Harper⁴⁷, Kristen M. Leraas⁴⁷, Lisa J. Wise⁴⁷, Teresa R. Tabler⁴⁷, Cynthia McAllister⁴⁷, Thomas Barr⁴⁷, Melissa Hart-Kothari⁴⁷

Tissue source sites: ABS-IUPUI Katie Tarvin⁴⁸, Charles Saller⁴⁹, George Sandusky⁵⁰, Colleen Mitchell⁵⁰, Christiana Mary V. Iacocca⁵¹, Jennifer Brown⁵¹, Brenda Rabeno⁵¹, Christine Czerwinski⁵¹, Nicholas Petrelli⁵¹; **Cureline** Oleg Dolzhansky⁵², Mikhail Abramov⁵³, Olga Voronina⁵⁴, Olga Potapova⁵⁴; **Duke University Medical Center** Jeffrey R. Marks⁵⁵; **The Greater Poland Cancer Center** Wiktor M. Suchorska⁵⁶, Dawid Murawa⁵⁶, Witold Kyrcer⁵⁶, Matthew Ibbs⁵⁶, Konstanty Kosiński⁵⁶, Arkadiusz Spychała⁵⁶, Paweł Murawa⁵⁶, Jacek J. Brzeziński⁵⁶, Hanna Perz⁵⁶, Radosław Łaźniak⁵⁶, Marek Teresiak⁵⁶, Honorata Tatka⁵⁶, Ewa Leporowska⁵⁶, Marta Bogusz-Czerniewicz^{56,57}, Julian Malicki^{56,57}, Andrzej Mackiewicz^{56,57}, Maciej Wiznerowicz^{56,57}; **ILSBio** Xuan Van Le⁵⁸, Bernard Kohl⁵⁸, Nguyen Viet Tien⁵⁹, Richard Thorp⁶⁰, Nguyen Van Bang⁶¹, Howard Sussman⁶², Bui Duc Phu⁶¹, Richard Hajek⁶³, Nguyen Phi Hung⁶⁴, Tran Viet The Phuong⁶⁵, Huynh Quyet Thang⁶⁶, Khurram Zaki Khan⁶⁰; **International Genomics Consortium** Robert Penny⁶⁷, David Mallery⁶⁷, Erin Curley⁶⁷, Candace Shelton⁶⁷, Peggy Yena⁶⁷; **Mayo Clinic** James N. Ingle⁶⁸, Fergus J. Couch⁶⁸, Wilma L. Lingle⁶⁸; **MSKCC** Tari A. King⁶⁹; **MD Anderson Cancer Center** Ana Maria Gonzalez-Angulo^{38,70}, Gordon B. Mills⁷⁰, Mary D. Dyer⁷⁰, Shuying Liu⁷⁰, Xiaolong Meng⁷⁰, Modesto Patangan⁷⁰; **University of California San Francisco** Frederic Waldman^{71,72}, Hubert Stöppler⁷³; **University of North Carolina** W. Kimryn Rathmell¹⁵, Leigh Thorne^{15,74}, Mei Huang^{15,74}, Lori Boice^{15,74}, Ashley Hill¹⁵; **Roswell Park Cancer Institute** Carl Morrison⁷⁵, Carmelo Gaudioso⁷⁵, Wiam Bshara⁷⁵; **University of Miami** Kelly Daily⁷⁶, Sophie C. Egea⁷⁶, Mark D. Pegram⁷⁶, Carmen Gomez-Fernandez⁷⁶; **University of Pittsburgh** Rajiv Dhir⁷⁷, Rohit Bhargava⁷⁸, Adam Brufsky⁷⁸; **Walter Reed National Military Medical Center** Craig D. Shriver⁷⁹, Jeffrey A. Hooke⁷⁹, Jamie Leigh Campbell⁷⁹, Richard J. Murali⁸⁰, Hai Hu⁸⁰, Stella Somari⁸⁰, Caroline Larson⁸⁰, Brenda Deyarmin⁸⁰, Leonid Brecher⁸⁰, Albert J. Kovatich⁸¹

Disease working group: Matthew J. Ellis^{3,82,83}, Tari A. King⁶⁹, Hai Hu⁸⁰, Fergus J. Couch⁶⁸, Richard J. Murali⁸⁰, Thomas Stricker⁸⁴, Kevin White⁸⁴, Olufunmilayo Olopade⁸⁵, James N. Ingle⁶⁸, Chunqing Luo⁸⁰, Yaqin Chen⁸⁰, Jeffrey R. Marks⁵⁵, Frederic Waldman^{71,72}, Maciej Wiznerowicz^{56,57}, Ron Bose^{3,82,83}, Li-Wei Chang⁸⁶, Andrew H. Beck¹⁰, Ana Maria Gonzalez-Angulo^{38,70}

Data coordination centre: Todd Pihl⁸⁷, Mark Jensen⁸⁷, Robert Sfeir⁸⁷, Ari Kahn⁸⁷, Anna Chu⁸⁷, Prachi Kothiyal⁸⁷, Zhining Wang⁸⁷, Eric Snyder⁸⁷, Joan Pontius⁸⁷, Brenda Ayala⁸⁷, Mark Backus⁸⁷, Jessica Walton⁸⁷, Julien Baboud⁸⁷, Dominique Berton⁸⁷, Matthew Nicholls⁸⁷, Deepak Srinivasan⁸⁷, Rohini Raman⁸⁷, Stanley Girshik⁸⁷, Peter Kigonya⁸⁷, Shelley Alonso⁸⁷, Rashmi Sanbhadri⁸⁷, Sean Barletta⁸⁷, David Pot⁸⁷

Project team: National Cancer Institute Margi Sheth⁸⁸, John A. Demchok⁸⁸, Kenna R. Mills Shaw⁸⁸, Liming Yang⁸⁸, Greg Eley⁸⁹, Martin L. Ferguson⁹⁰, Roy W. Tarnuzzer⁸⁸, Jianshan Zhang⁸⁸, Laura A. L. Dillon⁸⁸, Kenneth Buetow⁴⁴, Peter Fielding⁸⁸; **National Human Genome Research Institute** Bradley A. Ozenberger⁹¹, Mark S. Guyer⁹¹, Heidi J. Sofia⁹¹, Jacqueline D. Palchik⁹¹

¹The Genome Institute, Washington University, St Louis, Missouri 63108, USA.

²Department of Genetics, Washington University, St Louis, Missouri 63110, USA. ³Siteman Cancer Center, Washington University, St Louis, Missouri 63110, USA. ⁴Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia V5Z, Canada. ⁵The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.

⁶Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁷Department of Medicine, Harvard Medical School, Boston, Massachusetts 02215, USA. ⁸Departments of Cancer Biology and Medical Oncology, and the Center for

Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁹Department of Medical Oncology and the Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ¹⁰Department of Pathology, Harvard Medical School, Boston, Massachusetts 02215, USA. ¹¹Belfer Institute for Applied Cancer Science, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ¹²The Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹³Department of Genetics, Harvard Medical School and Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ¹⁴Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹⁵Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹⁶Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹⁷Institute for Pharmacogenetics and Individualized Therapy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹⁸Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹⁹Department of Internal Medicine, Division of Medical Oncology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ²⁰USC Epigenome Center, University of Southern California, Los Angeles, California 90033, USA. ²¹Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, Maryland 21231, USA. ²²Dan L Duncan Cancer Center, Baylor College of Medicine, Houston, Texas 77030, USA. ²³Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. ²⁴Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. ²⁵Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas 77030, USA. ²⁶The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02142, USA. ²⁷Institute for Applied Cancer Science, Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA. ²⁸Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA. ²⁹Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ³⁰Informatics Program, Children's Hospital, Boston, Massachusetts 02115, USA. ³¹Institute for Systems Biology, Seattle, Washington 98109, USA. ³²Tampere University of Technology, Tampere, Finland. ³³Cancer Genomics Core Laboratory, MD Anderson Cancer Center, Houston, Texas 77030, USA. ³⁴Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ³⁵Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ³⁶Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ³⁷Oregon Health and Science University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239, USA. ³⁸Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ³⁹Kleberg Center for Molecular Markers, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁴⁰Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁴¹Department of Biomolecular Engineering and Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA. ⁴²Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, California 95064, USA.

⁴³Buck Institute for Research on Aging, Novato, California 94945, USA. ⁴⁴Center for Bioinformatics and Information Technology, National Cancer Institute, Rockville, Maryland 20852, USA. ⁴⁵The Ohio State University College of Medicine, Department of Pathology, Columbus, Ohio 43205, USA. ⁴⁶The Ohio State University College of Medicine, Department of Pediatrics, Columbus, Ohio 43205, USA. ⁴⁷The Research Institute at Nationwide Children's Hospital, Columbus, Ohio 43205, USA. ⁴⁸ABS Inc. Indianapolis, Indiana 46204, USA. ⁴⁹ABS Inc. Wilmington, Delaware 19801, USA. ⁵⁰Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. ⁵¹Helen F. Graham Cancer Center, Christiana Care, Newark, Delaware 19713, USA. ⁵²Moscow City Clinical Oncology Dispensary 1 and the Central IHC Laboratory of the Moscow Health Department, Moscow 105005, Russia. ⁵³Russian Cancer Research Center, Moscow 115478, Russia. ⁵⁴Cureline, Inc., South San Francisco, California 94080, USA. ⁵⁵Department of Surgery, Duke University Medical Center, Durham, North Carolina 27710, USA. ⁵⁶The Greater Poland Cancer Centre, Poznań 61-866, Poland. ⁵⁷Poznan University of Medical Sciences, Poznań 61-701, Poland. ⁵⁸LSBio, LLC, Chestertown, Maryland 21620, USA. ⁵⁹Ministry of Health, Hanoi, Vietnam. ⁶⁰LSBio LLC, Karachi, Pakistan. ⁶¹Hue Central Hospital, Hue City, Vietnam. ⁶²Stanford University Medical Center, Stanford, California 94305, USA. ⁶³Center for Minority Health Research, University of Texas, MD Anderson Cancer Center, Houston, Texas 07703, USA. ⁶⁴National Cancer Institute, Hanoi, Vietnam. ⁶⁵Ho Chi Minh City Cancer Center, Vietnam. ⁶⁶Can Tho Cancer Center, Can Tho, Vietnam. ⁶⁷International Genomics Consortium, Phoenix, Arizona 85004, USA. ⁶⁸Mayo Clinic, Rochester, Minnesota 55905, USA. ⁶⁹Department of Surgery, Breast Service, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁷⁰Department of Breast Medical Oncology, The University of Texas, MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁷¹University of California at San Francisco; San Francisco, California 94143, USA. ⁷²Cancer Diagnostics; Nichols Institute, Quest Diagnostics; San Juan Capistrano, California 92675, USA. ⁷³Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, California 94115, USA. ⁷⁴UNC Tissue Procurement Facility, Department of Pathology, UNC Lineberger Cancer Center, Chapel Hill, North Carolina 27599, USA. ⁷⁵Department of Pathology, Roswell Park Cancer Institute, Buffalo, New York 14263, USA. ⁷⁶Department of Pathology, University of Miami Miller School of Medicine, Sylvester Comprehensive Cancer Center, Miami, Florida 33136, USA. ⁷⁷University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. ⁷⁸Magee-Womens Hospital of University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania 15213, USA. ⁷⁹Walter Reed National Military Medical Center, Bethesda, Maryland 20899-5600, USA. ⁸⁰Windber Research Institute, Windber, Pennsylvania 15963, USA. ⁸¹MDR Global, LLC, Windber, Pennsylvania 15963, USA. ⁸²Breast Cancer Program, Washington University, St Louis, Missouri 63110, USA. ⁸³Department of Internal Medicine, Division of Oncology, Washington University, St Louis, Missouri 63110, USA. ⁸⁴Institute for Genomics and Systems Biology, University of Chicago, Chicago, Illinois 60637, USA. ⁸⁵Center for Clinical Cancer Genetics, The University of Chicago, Chicago, Illinois 60637, USA. ⁸⁶Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ⁸⁷SRA International, 4300 Fair Lakes Court, Fairfax, Virginia 22033, USA. ⁸⁸The Cancer Genome Atlas Program Office, Center for Cancer Genomics, National Cancer Institute, Bethesda, Maryland 20852, USA. ⁸⁹TCGA Consultant, Scimientis, LLC, Statham, Georgia 30666, USA. ⁹⁰MLF Consulting, Arlington, Massachusetts 02474, USA. ⁹¹National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

Two stellar-mass black holes in the globular cluster M22

Jay Strader^{1,2}, Laura Chomiuk^{1,2,3}, Thomas J. Maccarone⁴, James C. A. Miller-Jones⁵ & Anil C. Seth⁶

Hundreds of stellar-mass black holes probably form in a typical globular star cluster, with all but one predicted to be ejected through dynamical interactions^{1–3}. Some observational support for this idea is provided by the lack of X-ray-emitting binary stars comprising one black hole and one other star ('black-hole/X-ray binaries') in Milky Way globular clusters, even though many neutron-star/X-ray binaries are known⁴. Although a few black holes have been seen in globular clusters around other galaxies^{5,6}, the masses of these cannot be determined, and some may be intermediate-mass black holes that form through exotic mechanisms⁷. Here we report the presence of two flat-spectrum radio sources in the Milky Way globular cluster M22, and we argue that these objects are black holes of stellar mass (each ~ 10 – 20 times more massive than the Sun) that are accreting matter. We find a high ratio of radio-to-X-ray flux for these black holes, consistent with the larger predicted masses of black holes in globular clusters compared to those outside⁸. The identification of two black holes in one cluster shows that ejection of black holes is not as efficient as predicted by most models^{1,2,4}, and we argue that M22 may contain a total population of ~ 5 – 100 black holes. The large core radius of M22 could arise from heating produced by the black holes⁹.

We have obtained very deep radio continuum images of the Milky Way globular cluster M22 (NGC 6656) with the Karl G. Jansky Very Large Array (VLA). The principal goal of the observations was to search for a possible central intermediate-mass black hole via synchrotron emission from the accretion of intracluster gas; no central source was found¹⁰. However, we serendipitously detected two previously unknown radio continuum sources in the core of the cluster (Fig. 1). We term the sources M22-VLA1 and M22-VLA2. Both sources have flat radio spectra and are unresolved at our $\sim 1''$ resolution.

The core radius of M22 is uncommonly large for a Milky Way globular cluster, namely, ~ 1.24 pc (ref. 11). These sources are well inside the cluster core, at projected radii of 0.4 pc and 0.25 pc for M22-VLA1 and M22-VLA2, respectively. The next source of comparable flux density is far outside the core, at a projected radius of 2.4 pc. These sources have no counterparts in shallow archival Chandra X-ray imaging. On the basis of these non-detections, the sources are constrained to have X-ray luminosities $L_X \lesssim 2.2 \times 10^{30}$ erg s⁻¹ over 3–9 keV at the distance of M22¹². The radio luminosities of the sources at 8.4 GHz are $L_R \approx 6 \times 10^{27}$ erg s⁻¹, assuming flat spectra. Therefore, if the sources are not variable, the limit of radio to X-ray luminosity is $\log(L_R/L_X) \gtrsim -2.6$.

The radio luminosity, L_R/L_X ratio and central location of the sources place significant constraints on their nature. The most likely explanation is that both sources are accreting stellar-mass black holes in M22. Other possibilities, all of which we consider unlikely, are discussed in Supplementary Information. These objects are the first strong candidates for stellar-mass black holes in any Milky Way globular cluster, and the first stellar-mass black holes to be discovered through radio emission rather than via X-rays¹³.

The radio emission implies that the black holes are actively accreting, and the flat radio spectra are consistent with relatively low accretion rates¹⁴ ($\lesssim 2$ – 3% of the Eddington rate). Because globular clusters have modest amounts of interstellar gas, it is very unlikely that the radio luminosity can be explained by Bondi accretion. Thus the objects cannot be black-hole/black-hole binaries, and instead are probably in binary systems with Roche lobe-overflowing companions. Stellar-mass black holes, ~ 5 – 100 times the solar mass, M_\odot , offer the best explanation for the presence of multiple sources close to the cluster centre; objects more massive than the average cluster star will sink to the centre because of mass segregation.

To look for optical counterparts of the radio sources, we used archival Hubble Space Telescope (HST) imaging of M22, for which photometric catalogues are available¹⁵. Figure 2 shows that M22-VLA1 is a close match ($0.05''$) to a moderately low-mass ($\sim 0.34 M_\odot$) main sequence M dwarf in M22, as inferred from standard stellar isochrones (see Supplementary Information for more details). M22-VLA2 is $0.17''$

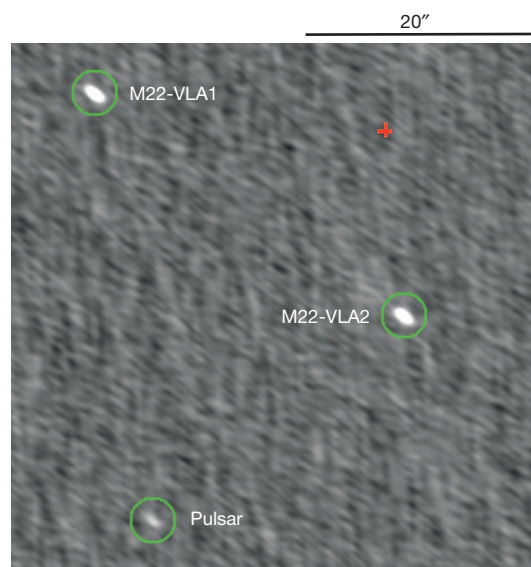


Figure 1 | VLA radio continuum image of the core of the globular cluster M22. The two bright circled objects are the sources identified as stellar-mass black holes, M22-VLA1 and M22-VLA2. These sources have flux densities of 55–58 μ Jy at 5.9 GHz. We obtained the data in two separate 1-GHz-wide basebands centred at 5 and 6.75 GHz, allowing a measurement of the spectral index of the radio emission between these frequencies. Both sources have flat radio spectra, with $\alpha = 0$ – 0.2 , assuming $S_\nu = \nu^\alpha$ (here S_ν is flux density, and ν is frequency). The faint circled object is a known millisecond pulsar²⁷. A red cross marks the photometric cluster centre. $20''$ corresponds to approximately 0.3 pc at the distance of M22. The apparent elongation of the sources is due entirely to the elongated synthesized beam; all three circled sources are unresolved. North is up and east is to the left in this image.

¹Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA. ²Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA. ³National Radio Astronomy Observatory, PO Box O, Socorro, New Mexico 87801, USA. ⁴School of Physics and Astronomy, University of Southampton, Highfield SO17 1BJ, UK. ⁵International Centre for Radio Astronomy Research, Curtin University, GPO Box U1987, Perth, Western Australia 6845, Australia. ⁶Department of Physics and Astronomy, University of Utah, Salt Lake City, Utah 84112, USA.

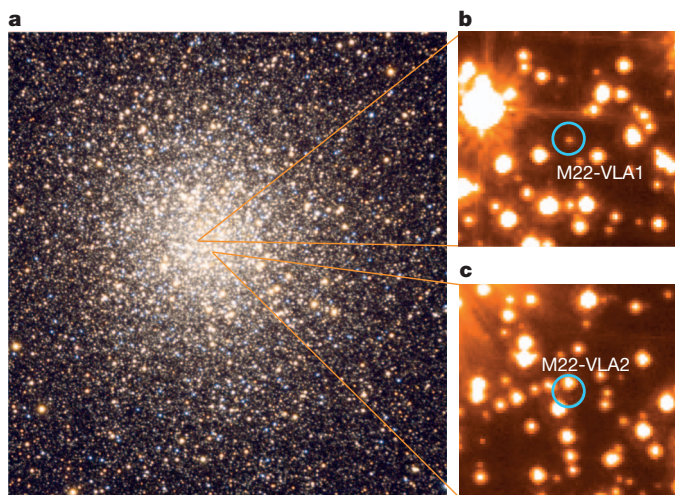


Figure 2 | Optical images of M22 and the candidate companion stars to the radio sources. **a**, Ground-based image that shows the approximate location of the sources in the context of the star cluster. **b**, **c**, The zoomed-in location of the radio sources (**b**, M22-VLA1; **c**, M22-VLA2) on an archival high-resolution HST/Advanced Camera for Surveys F814W image. Each blue circle has a radius of $0.3''$ for clarity; the uncertainty in the astrometric matching of the optical and radio data is $<0.1''$. The image orientation is as in Fig. 1. (Image credit for **a**: D. Matthews/A. Block/NOAO/AURA/NSF.)

from the nearest detected star, which is a $\sim 0.62 M_{\odot}$ main sequence star. Considering the distribution of stars in the inner $30''$ of the cluster, the probability of a chance coincidence as close as for M22-VLA1 is only 2%; for M22-VLA2 it is 26%. Thus we consider the optical association for source M22-VLA1 suggestive, but that for M22-VLA2 uncertain. However, for the case of M22-VLA1, there is an additional complication: because the average stellar mass in the core is greater than that of the putative companion, the low-mass main sequence star would probably be exchanged out of the binary in a three-body interaction with another star⁴. On the other hand, because of the low central density of M22¹¹ ($<10^4 M_{\odot} \text{ pc}^{-3}$), a binary with a low-mass companion might survive longer than in a typical globular cluster. Nonetheless, it is possible that both radio sources are associated with low-luminosity objects below the detection limit of the HST data, such as white dwarfs.

Stellar-mass black holes with accretion rates below $\sim 2\%$ of the Eddington rate¹⁶ (in the so-called low/hard state) follow an empirical correlation between radio and X-ray luminosity with a scatter of a factor of about two (ref. 17). Figure 3 shows this correlation with the M22 data overplotted. The radio–X-ray relation predicts an X-ray luminosity of 10^{31} – $10^{32} \text{ erg s}^{-1}$ for this radio luminosity^{18,19}, above the completeness limit of the archival Chandra data. There are several plausible explanations for this discrepancy. First, there is the possibility of variability. The X-ray data were taken in 2005, six years earlier than the radio data. Field stellar-mass black holes in the low/hard state show substantial (typically a factor of 2–10) variability in both radio and X-rays^{20,21}. Therefore, concurrent radio and X-ray data are necessary for precise constraints on L_R/L_X . We found marginal evidence for radio variability in M22-VLA2 on the timescale of a week; more details can be found in Supplementary Information. Another plausible explanation is that there is larger scatter in the radio–X-ray correlation at very low accretion rates. Only a single known black-hole binary has a measured radio luminosity as faint as our sources¹⁸, and there is evidence that some stellar-mass black holes with low X-ray luminosities may not fall on the correlation¹⁹.

An intriguing possibility is that these sources have high values of L_R/L_X because they are more massive than typical stellar-mass black holes in the field. The radio–X-ray correlation for stellar-mass black holes is a special case of a ‘fundamental plane’ for black-hole accretion in the low/hard state that includes the black-hole mass as a third

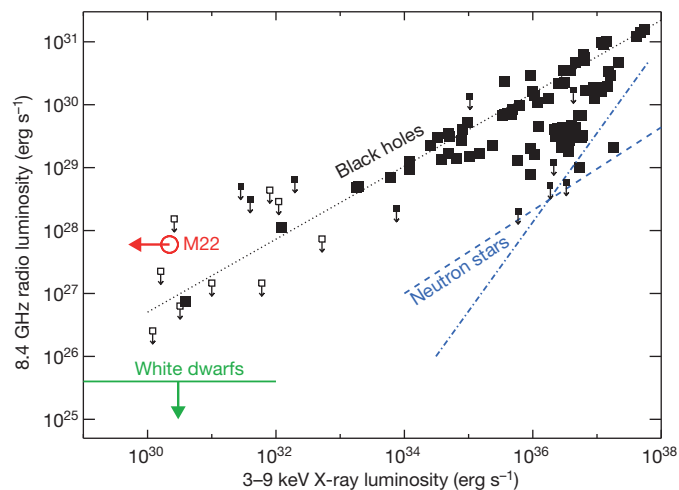


Figure 3 | Radio–X-ray correlation for stellar-mass black holes. The M22 sources have properties more consistent with black holes than with neutron stars or white dwarfs. Filled squares represent simultaneous radio and X-ray data; open squares are non-simultaneous measurements, the positions of which might have been affected by variability. Upper limits are also shown. Some objects have multiple measurements plotted that represent different phases of accretion. The open red circle represents both M22-VLA1 and M22-VLA2, which have very similar luminosities. The dotted black line represents the published correlation¹⁸ $L_R \propto L_X^{0.58}$, normalized by a least-squares fit to the simultaneous detections with $L_X < 2 \times 10^{34} \text{ erg s}^{-1}$. The dashed and dot-dashed blue lines show two possible radio–X-ray correlations for accreting neutron stars; this relation is poorly constrained by observations²⁸. The solid green line shows the maximum radio continuum luminosity observed for accreting white dwarfs²⁹. Neither neutron stars nor white dwarfs have properties consistent with the M22 radio sources. More information can be found in Supplementary Information.

parameter²². In this relation, more massive black holes have larger values of L_R/L_X . If our sources have masses of ~ 15 – $20 M_{\odot}$ rather than the 5 – $10 M_{\odot}$ typical of field stellar-mass black holes²³, then their X-ray luminosities should be lower than predicted by the correlation in Fig. 3 by a factor of ~ 2 – 3 . It is reasonable to expect that black holes in globular clusters will be more massive than those in the field. Field black holes with measured dynamical masses are all in binary systems, and were probably affected by mass transfer during a common envelope stage that reduced the mass of the resulting black holes²⁴. This need not be the case in globular clusters, because black holes can form as single objects or in wide binaries, and then be exchanged into pre-existing binaries or tidally capture companions owing to the high stellar densities²⁵. Globular cluster black holes also form at lower metallicity than in the field, leading to less mass loss from the progenitor and thus more massive remnants⁸.

As mentioned above, the location of stars in a cluster also gives information about their masses. Stellar-mass black holes will mass-segregate to the core of the cluster. This process can be used to roughly estimate their masses by assuming thermalization, for which this relation holds¹: $m_{\text{BH}}/m_{\star} = (r_c/r_{\text{BH}})^2$, where m_{BH} and r_{BH} are the characteristic black-hole mass and radius, m_{\star} is the typical stellar mass, and r_c is the core radius. Assuming $m_{\star} = 1 M_{\odot}$ in the segregated cluster core and taking the observed values of $r_c = 1.24 \text{ pc}$ and $r_{\text{BH}} = 0.33 \text{ pc}$, we estimate $m_{\text{BH}} \approx 15 M_{\odot}$.

The existence of black holes in a low-density globular cluster such as M22 constrains the magnitude of the initial velocity kicks received by the black holes at birth. The current central escape velocity of M22¹¹ is $\sim 34 \text{ km s}^{-1}$. This value may have been higher in the past, owing to a larger cluster mass and a more compact structure. Nonetheless, the retention of two black holes in a globular cluster with a modest escape velocity implies that the black holes could not have received large natal kicks. Large kicks are inferred for some stellar-mass black holes in the field²⁶. Low kick velocities can originate from supernovae if the

black-hole mass is large, or if the black holes form from direct collapse with no supernovae. In either case, higher black-hole masses are favoured.

The presence of black holes in a globular cluster can lead to an expansion of the core radius through interactions between black holes and stars. This could explain why M22 has the fifth-largest core radius among luminous ($\geq 2 \times 10^5 L_{\odot}$) Milky Way globular clusters¹¹. Additional discussion can be found in Supplementary Information. Most theoretical models in the literature predict that only a single black hole (or black-hole/black-hole binary) will survive the dynamical processes by which black holes mass-segregate to the cluster centre, form an unstable subcluster, and evaporate^{1,2,4}. In some cases, more than one black hole may temporarily survive for an additional black-hole relaxation time (< 1 Gyr), if the extra black holes are kicked into orbits outside the core². Additional discussion can be found in Supplementary Information.

In contrast to these^{1,2,4} theoretical predictions, M22 contains more than one black hole. In fact, it is possible that more than two black holes are present in M22, either as single black holes or in binary systems that are not undergoing observable mass transfer. Under the uncertain assumption that both of the M22 sources are black-hole/white-dwarf binaries, published calculations can be used to estimate the fraction of surviving black holes that are actively accreting in present-day globular clusters²⁵. Over 10 Gyr, 2–40% of black holes are expected to become members of binary systems with observable accretion. Our two observed sources thus suggest a total population of ~ 5 –100 black holes in M22.

Received 19 April; accepted 10 August 2012.

- Kulkarni, S. R., Hut, P. & McMillan, S. Stellar black holes in globular clusters. *Nature* **364**, 421–423 (1993).
- Sigurdsson, S. & Hernquist, L. Primordial black holes in globular clusters. *Nature* **364**, 423–425 (1993).
- Portegies Zwart, S. F. & McMillan, S. L. W. Black hole mergers in the universe. *Astrophys. J.* **528**, L17–L20 (2000).
- Kalogera, V., King, A. R. & Rasio, F. A. Could black hole X-ray binaries be detected in globular clusters? *Astrophys. J.* **601**, L171–L174 (2004).
- Maccarone, T. J., Kundu, A., Zepf, S. E. & Rhode, K. L. A black hole in a globular cluster. *Nature* **445**, 183–185 (2007).
- Irwin, J. A., Brink, T. G., Bregman, J. N. & Roberts, T. P. Evidence for a stellar disruption by an intermediate-mass black hole in an extragalactic globular cluster. *Astrophys. J.* **712**, L1–L4 (2010).
- Portegies Zwart, S. F., Baumgardt, H., Hut, P., Makino, J. & McMillan, S. L. W. Formation of massive black holes through runaway collisions in dense young star clusters. *Nature* **428**, 724–726 (2004).
- Belczynski, K. *et al.* On the maximum mass of stellar black holes. *Astrophys. J.* **714**, 1217–1226 (2010).
- Mackey, A. D., Wilkinson, M. I., Davies, M. B. & Gilmore, G. F. Black holes and core expansion in massive star clusters. *Mon. Not. R. Astron. Soc.* **386**, 65–95 (2008).
- Strader, J. *et al.* No evidence for intermediate-mass black holes in globular clusters: strong constraints from the VLA. *Astrophys. J.* **750**, L27 (2012).
- McLaughlin, D. & van der Marel, R. Resolved massive star clusters in the Milky Way and its satellites: brightness profiles and a catalog of fundamental parameters. *Astrophys. J. Suppl. Ser.* **161**, 304–360 (2005).
- Monaco, L., Pancino, E., Ferraro, F. R. & Bellazzini, M. Wide-field photometry of the Galactic globular cluster M22. *Mon. Not. R. Astron. Soc.* **349**, 1278–1290 (2004).
- Maccarone, T. & Knigge, C. Compact objects in globular clusters. *Astron. Geophys.* **48**, 5.12–5.20 (2007).
- Gallo, E., Fender, R. P. & Hynes, R. I. The radio spectrum of a quiescent stellar mass black hole. *Mon. Not. R. Astron. Soc.* **356**, 1017–1021 (2005).
- Anderson, J. *et al.* The ACS survey of Galactic globular clusters. V. Generating a comprehensive star catalog for each cluster. *Astron. J.* **135**, 2055–2073 (2008).
- Maccarone, T. J. Do X-ray binary spectral state transition luminosities vary? *Astron. Astrophys.* **409**, 697–706 (2003).
- Gallo, E., Fender, R. P. & Pooley, G. G. A universal radio-X-ray correlation in low/hard state black hole binaries. *Mon. Not. R. Astron. Soc.* **344**, 60–72 (2003).
- Gallo, E. *et al.* A radio-emitting outflow in the quiescent state of A0620–00: implications for modelling low-luminosity black hole binaries. *Mon. Not. R. Astron. Soc.* **370**, 1351–1360 (2006).
- Miller-Jones, J. C. A., Jonker, P. G., Maccarone, T. J., Nelemans, G. & Calvelo, D. E. A deep radio survey of hard state and quiescent black hole X-Ray binaries. *Astrophys. J.* **739**, L18 (2011).
- Miller-Jones, J. C. A. *et al.* Zooming in on a sleeping giant: milliarcsecond High Sensitivity Array imaging of the black hole binary V404 Cyg in quiescence. *Mon. Not. R. Astron. Soc.* **388**, 1751–1758 (2008).
- Corbel, S., Tomsick, J. A. & Kaaret, P. On the origin of black hole X-Ray emission in quiescence: Chandra observations of XTE J1550–564 and H1743–322. *Astrophys. J.* **636**, 971–978 (2006).
- Merloni, A., Heinz, S. & di Matteo, T. A fundamental plane of black hole activity. *Mon. Not. R. Astron. Soc.* **345**, 1057–1076 (2003).
- Remillard, R. A. & McClintock, J. E. X-ray properties of black-hole binaries. *Annu. Rev. Astron. Astrophys.* **44**, 49–92 (2006).
- Tauris, T. M. & van den Heuvel, E. P. J. in *Compact Stellar X-ray Sources* (eds Lewin, W. H. G. & van der Klis, M.) 623–665 (Cambridge Univ. Press, 2006).
- Ivanova, N. *et al.* Formation of black hole X-ray binaries in globular clusters. *Astrophys. J.* **717**, 948–957 (2010).
- Repetto, S., Davies, M. B. & Sigurdsson, S. Investigating stellar-mass black hole kicks. *Mon. Not. R. Astron. Soc.* (in the press); preprint at <http://arXiv.org/abs/1203.3077> (2012).
- Lynch, R. S., Ransom, S. M., Freire, P. C. C. & Stairs, I. H. Six new recycled globular cluster pulsars discovered with the Green Bank Telescope. *Astrophys. J.* **734**, 89 (2011).
- Migliari, S. & Fender, R. P. Jets in neutron star X-ray binaries: a comparison with black holes. *Mon. Not. R. Astron. Soc.* **366**, 79–91 (2006).
- Körding, E. G., Knigge, C., Tzioumis, T. & Fender, R. Detection of radio emission from a nova-like cataclysmic variable: evidence of jets? *Mon. Not. R. Astron. Soc.* **418**, L129–L132 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. L.C. is a Jansky Fellow of the National Radio Astronomy Observatory. This work is partially based on observations made with the NASA/ESA Hubble Space Telescope, and obtained from the Hubble Legacy Archive, which is a collaboration between the Space Telescope Science Institute (STScI/NASA), the Space Telescope European Coordinating Facility (ST-ECF/ESA) and the Canadian Astronomy Data Centre (CAD/C/NRC/CSA).

Author Contributions J.S. wrote the text. L.C. reduced the data. All authors contributed to the interpretation of the data and commented on the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.S. (strader@pa.msu.edu).

Comet-like mineralogy of olivine crystals in an extrasolar proto-Kuiper belt

B. L. de Vries¹, B. Acke¹, J. A. D. L. Blommaert¹, C. Waelkens¹, L. B. F. M. Waters^{2,3}, B. Vandenbussche¹, M. Min³, G. Olofsson⁴, C. Dominik^{3,5}, L. Decin^{1,3}, M. J. Barlow⁶, A. Brandeker⁴, J. Di Francesco⁷, A. M. Glauser^{8,9}, J. Greaves¹⁰, P. M. Harvey¹¹, W. S. Holland^{9,12}, R. J. Ivison⁹, R. Liseau¹³, E. E. Pantin¹⁴, G. L. Pilbratt¹⁵, P. Royer¹ & B. Sibthorpe⁹

Some planetary systems harbour debris disks containing planetesimals such as asteroids and comets¹. Collisions between such bodies produce small dust particles², the spectral features of which reveal their composition and, hence, that of their parent bodies. A measurement of the composition of olivine crystals ($\text{Mg}_{2-2x}\text{Fe}_{2x}\text{SiO}_4$) has been done for the protoplanetary disk HD 100546 (refs 3, 4) and for olivine crystals in the warm inner parts of planetary systems. The latter compares well with the iron-rich olivine in asteroids^{5,6} ($x \approx 0.29$). In the cold outskirts of the β Pictoris system, an analogue to the young Solar System, olivine crystals were detected⁷ but their composition remained undetermined, leaving unknown how the composition of the bulk of Solar System cometary olivine grains compares with that of extrasolar comets^{8,9}. Here we report the detection of the 69-micrometre-wavelength band of olivine crystals in the spectrum of β Pictoris. Because the disk is optically thin, we can associate the crystals with an extrasolar proto-Kuiper belt a distance of 15–45 astronomical units from the star (one astronomical unit is the Sun–Earth distance), determine their magnesium-rich composition ($x = 0.01 \pm 0.001$) and show that they make up 3.6 ± 1.0 per cent of the total dust mass. These values are strikingly similar to those for the dust emitted by the most primitive comets in the Solar System^{8–10}, even though β Pictoris is more massive and more luminous and has a different planetary system architecture.

The olivine crystals found in the Itokawa asteroid and in ordinary chondrites (types 4 to 6) have an iron-rich composition⁵ ($x \approx 0.29$). In contrast, laboratory measurements of olivine crystals from unequilibrium bodies such as comet 81P/Wild 2 and cometary interplanetary dust particles show that these crystals have a range of compositions, but the distribution has a pronounced and sharp peak at the almost-pure magnesium-rich composition with $x \approx 0.01$ (refs 8, 9). Both laboratory experiments¹¹ and observations^{3,4} show that crystal formation in protoplanetary disks by gas-phase condensation, thermal annealing and shock heating results in magnesium-rich crystalline olivine^{12–16} ($x < 0.1$). During the protoplanetary disk phase, these olivine crystals are incorporated into planetesimals. An example of a planetary system in which the olivine crystals are then freed from such planetesimals by collisions is the system of β Pictoris. This system is a young (~ 12 Myr) analogue to the Solar System, with at least one planet at a distance of ~ 10 AU and a dusty debris disk containing small dust grains^{7,17–20} (Fig. 1).

We have detected (Fig. 1) the 69- μm spectral band of small (~ 2 - μm ; see Supplementary Information), crystalline olivine grains in the planetary system of β Pictoris using Herschel²¹ PACS²². The 69- μm band is of special interest because its exact peak wavelength

and width are sensitive to both the grain temperature and, in particular, the composition of the olivine crystals^{23,24} (Fig. 2). From our model fitting of the 69- μm band and spectral bands at shorter wavelengths (Fig. 1), the temperature and total mass of the crystals are determined

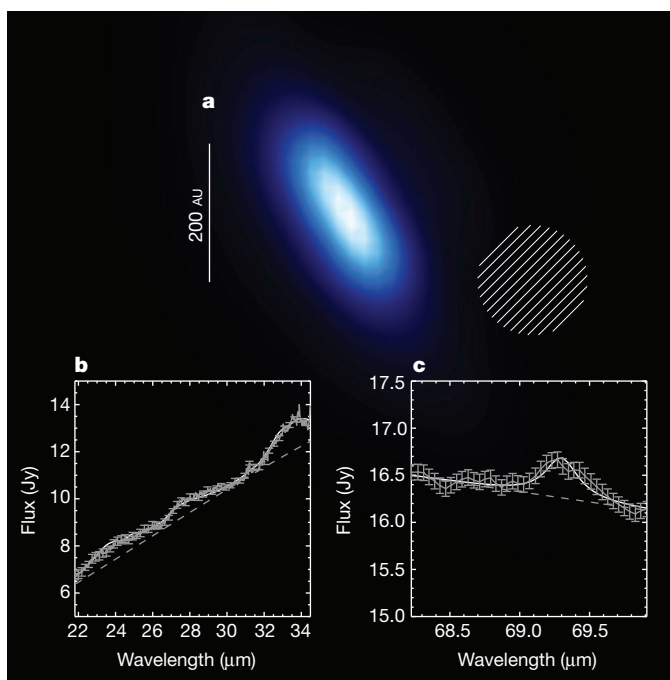


Figure 1 | Photometric and spectral observations of the planetary system of β Pictoris. **a**, Resolved surface brightness map of the β Pictoris debris disk at 70 μm taken with the Herschel Space Observatory's²¹ Photodetector Array Camera and Spectrometer²² (PACS). The disk is barely resolved with PACS, which has a point-spread function with a full-width at half-maximum of $8.2''$ (hatched circle). **b**, Spitzer Space Telescope infrared spectrograph spectrum showing prominent olivine features⁷ (solid grey). The white solid line is our best model fit and the grey dashed line is the continuum. The uncertainties (1σ) in the Spitzer data⁷ are indicated in the figure. **c**, The flux-corrected PACS spectrum with error bars (1σ) showing the 69- μm band of crystalline olivine (solid grey; 12σ detection). The white solid line shows the model fit to the 69- μm band of crystalline olivine as described in Supplementary Information, and the dashed grey line shows the underlying dust continuum. The best model contains crystalline olivine ($\text{Mg}_{2-2x}\text{Fe}_{2x}\text{SiO}_4$) with $x = 0.01 \pm 0.001$ (1σ) and a temperature of 85 ± 6 K (1σ).

¹Instituut voor Sterrenkunde, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium. ²SRON, Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands. ³Astronomical Institute "Anton Pannekoek", University of Amsterdam, PO Box 94249, 1090 GE Amsterdam, The Netherlands. ⁴Department of Astronomy, Stockholm University, AlbaNova University Center, 106 91 Stockholm, Sweden. ⁵Department of Astrophysics/IMAPP, Radboud University Nijmegen, PO Box 9010, 6500 GL Nijmegen, The Netherlands. ⁶Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK. ⁷National Research Council of Canada, 5071 West Saanich Road, Victoria, British Columbia, V9E 2E7, Canada. ⁸ETH Zurich, Institute for Astronomy, Wolfgang-Pauli-Strasse 27, 8093 Zurich, Switzerland. ⁹UK Astronomy Technology Centre, Royal Observatory Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK. ¹⁰SUPA, Physics and Astronomy, North Haugh, St Andrews, Fife KY16 9SS, UK. ¹¹Astronomy Department, University of Texas, Austin, Texas 78712, USA. ¹²Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, UK. ¹³Earth and Space Sciences, Chalmers University of Technology, Onsala Space Observatory, 439 92 Onsala, Sweden. ¹⁴Laboratoire AIM, CEA/DSM-CNRS-Université Paris Diderot, IRFU/Service d'Astrophysique, Bâtiment 709, CEA-Saclay, 91191 Gif-sur-Yvette Cedex, France. ¹⁵ESA Research and Scientific Support Department, ESTEC/SRE-SA, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands.

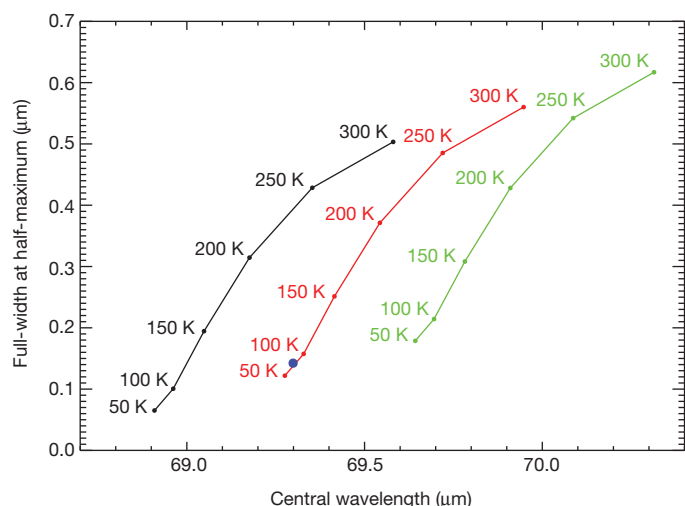


Figure 2 | Diagram demonstrating the dependence of the 69- μ m band on grain temperature and composition. The diagram gives the width and central wavelength of the 69- μ m band for six temperatures and for crystalline olivine ($\text{Mg}_{2-2x}\text{Fe}_{2x}\text{SiO}_4$) with $x = 0.0$ (black), $x = 0.01$ (red) and $x = 0.02$ (green). The width and central wavelength are measured by fitting Lorentzian profiles to laboratory measurements^{23,24} of crystalline olivine at different temperatures and compositions (see Supplementary Information for additional information). The width and wavelength positions measured show how the band broadens and shifts as a function of temperature or iron content. The best model fit of the 69- μ m band of β Pictoris is indicated with a solid blue dot (Fig. 1b, c, white solid line); that is, the olivine crystals are cold (85 ± 6 K) and contain about 1% iron ($x = 0.01 \pm 0.001$).

to be 85 ± 6 K and $(2.8 \pm 0.8) \times 10^{23}$ g, respectively. The exact wavelength position of the 69- μ m band indicates very magnesium-rich crystalline olivine ($x = 0.01 \pm 0.001$ (1σ)). The fraction of olivine crystals to the total amount of dust (obtained from the spectral energy distribution; see Supplementary Information) is $3.6 \pm 1.0\%$ (1σ). The temperature of 85 ± 6 K (1σ) places the population of olivine crystals between 15 and 45 AU from the central star, which coincides with a strong increase in surface density in the disk²⁵. This location is outside the snow line of the system, where icy, comet-like bodies can exist, such as in the Kuiper belt of the Solar System. Scaling the distances in the β Pictoris system to those of the Solar System according to the different luminosities of the two central stars, the extrasolar Kuiper belt of β Pictoris reaches into the temperature range of the Jupiter–Saturn region. We propose that this location is an inward extension of what will in time become an analogue of the Kuiper belt of the Solar System.

The composition of the olivine crystals around β Pictoris is strikingly similar to that found in cometary bodies in the Solar System. From the low iron content, we can conclude that the olivine crystals we observe in β Pictoris come from collisions between unequilibrated, relatively small (<10 -km) comet-like bodies⁵. The magnesium-rich olivine crystals around β Pictoris are in stark contrast to the iron-rich crystalline olivine⁵ ($x = 0.29$) found in asteroid-like bodies in the Solar System. When we compare the crystalline olivine abundance found in β Pictoris ($3.6 \pm 1.0\%$) to that of primitive comets in the Solar System, similar low values are found. The comets 17P/Holmes and 73P/Schwassmann–Wachmann, for example, contain about ~ 2 – 10% crystalline olivine compared with the total amount of dust^{10,26,27}. Because olivine crystals can be formed only within 10 AU of the central star^{12–15} there must have been a transportation mechanism to bring these crystals to Kuiper belt distances. Studies of crystalline material and gas have indeed shown that radial mixing has taken place in both the Solar System and disks around young stars^{28,29}. Models are able to predict crystalline olivine abundances of 2–58% at radii beyond 10 AU on timescales of ~ 1 Myr (refs 12, 30). The similar crystalline olivine abundances in β Pictoris and Solar System comets suggest that radial

mixing must have been at work during the formation of the β Pictoris planetary system, with an efficiency similar to that in the protosolar nebula.

Received 21 December 2011; accepted 6 August 2012.

- Calvet, N. *et al.* Disks in transition in the Taurus population: Spitzer IRS spectra of GM Aurigae and DM Tauri. *Astrophys. J.* **630**, L185 (2005).
- Wyatt, M. C. *et al.* Steady state evolution of debris disks around A stars. *Astrophys. J.* **663**, 365–382 (2007).
- Sturm, B. *et al.* First results of the Herschel key program “Dust, Ice and Gas In Time” (DIGIT): dust and gas spectroscopy of HD 100546. *Astron. Astrophys.* **518**, L129 (2010).
- Mulders, G. *et al.* Low abundance, strong features: window-dressing crystalline forsterite in the disk wall of HD 100546. *Astron. Astrophys.* **531**, A93 (2011).
- Nakamura, T. *et al.* Itokawa dust particles: a direct link between S-type asteroids and ordinary chondrites. *Science* **333**, 1113–1116 (2011).
- Olofsson, J. *et al.* Transient dust in warm debris disks. *Astron. Astrophys.* **542**, A90 (2012).
- Chen, C. H. *et al.* The dust and gas around β Pictoris. *Astrophys. J.* **666**, 466–474 (2007).
- Zolensky, M. E. *et al.* Mineralogy and petrology of comet 81P/Wild 2 nucleus samples. *Science* **314**, 1735–1739 (2006).
- Zolensky, M. E. *et al.* Comparing Wild 2 particles to chondrites and IDPs. *Meteorit. Planet. Sci.* **43**, 261–272 (2008).
- Sitko, M. L. *et al.* Infrared spectroscopy of comet 73P/Schwassmann–Wachmann 3 using the Spitzer Space Telescope. *Astron. J.* **142**, 80–89 (2011).
- Davoine, C. *et al.* The origin of GEMS in IDPs as deduced from microstructural evolution of amorphous silicates with annealing. *Astron. Astrophys.* **448**, L1 (2006).
- Gail, H. P. Radial mixing in protoplanetary accretion disks. IV. Metamorphosis of the silicate dust complex. *Astron. Astrophys.* **413**, 571–591 (2004).
- Nuth, J. A. & Johnson, N. M. Crystalline silicates in comets: how did they form? *Icarus* **180**, 243–250 (2006).
- Tanaka, K. K., Yamamoto, T. & Kimura, H. Low-temperature crystallization of amorphous silicate in astrophysical environments. *Astrophys. J.* **717**, 586–596 (2010).
- Abrahám, P. *et al.* Episodic formation of cometary material in the outburst of a young Sun-like star. *Nature* **459**, 224–226 (2009).
- Kemper, F., Vriend, W. J. & Tielens, A. G. G. M. The absence of crystalline silicates in the diffuse interstellar medium. *Astrophys. J.* **609**, 826–837 (2004).
- Lecavelier des Etangs, A. *et al.* Deficiency of molecular hydrogen in the disk of β Pictoris. *Nature* **412**, 706–708 (2001).
- Okamoto, Y. K. *et al.* An early extrasolar planetary system revealed by planetesimal belts in β Pictoris. *Nature* **431**, 660–663 (2004).
- Lagrange, A. M. *et al.* A giant planet imaged in the disk of the young star β Pictoris. *Science* **329**, 57–59 (2010).
- Vandenbusche, B. *et al.* The β Pictoris disk imaged by Herschel PACS and SPIRE. *Astron. Astrophys.* **518**, L133 (2010).
- Pilbratt, G. L. *et al.* Herschel Space Observatory: an ESA facility for far-infrared and submillimetre astronomy. *Astron. Astrophys.* **518**, L1 (2010).
- Poglitsch, A. *et al.* The Photodetector Array Camera and Spectrometer (PACS) on the Herschel Space Observatory. *Astron. Astrophys.* **518**, L2 (2010).
- Koike, C. *et al.* Compositional dependence of infrared absorption spectra of crystalline silicate. II. Natural and synthetic olivines. *Astron. Astrophys.* **399**, 1101–1107 (2003).
- Suto, H. *et al.* Low-temperature single crystal reflection spectra of forsterite. *Mon. Not. R. Astron. Soc.* **370**, 1599–1606 (2006).
- Lagage, P. O. & Pantin, E. Dust depletion in the inner disk of β Pictoris as a possible indicator of planets. *Nature* **369**, 628–630 (1994).
- Reach, W. T. *et al.* Distribution and properties of fragments and debris from the split Comet 73P/Schwassmann–Wachmann 3 as revealed by Spitzer Space Telescope. *Icarus* **203**, 571–588 (2009).
- Lisse, C. M. *et al.* On the nature of the dust in the debris disk around HD 69830. *Astrophys. J.* **658**, 584–592 (2007).
- van Boekel, R. *et al.* The building blocks of planets within the ‘terrestrial’ region of protoplanetary disks. *Nature* **432**, 479–482 (2004).
- Hartogh, P. *et al.* Ocean-like water in the Jupiter-family comet 103P/Hartley 2. *Nature* **478**, 218–220 (2011).
- Vinković, D. Radiation-pressure mixing of large dust grains in protoplanetary disks. *Nature* **459**, 227–229 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements Herschel is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA. PACS has been developed by a consortium of institutes led by MPE (Germany) and including UVIE (Austria); KUL, CSL and IMEC (Belgium); CEA and OAMP (France); MPIA (Germany); IFSI, OAP/AOT, OAA/CAISMI, LENS and SISSA (Italy); and IAC (Spain). This development has been supported by the funding agencies BMVIT (Austria), ESA-PRODEX (Belgium), CEA/CNES (France), DLR (Germany), ASI (Italy) and CICT/MCT (Spain). B.L.d.V. is an Aspirant Fellow of the Fund for Scientific Research, Flanders.

Author Contributions B.L.d.V. analysed the data, performed the modelling and wrote the manuscript; B.A. was heavily involved in designing the study, doing the analysis,

applying the statistical methods and in writing the paper; J.A.D.L.B. did the data reduction and helped in writing the manuscript; C.W. was largely responsible for the study design and obtaining the observations, and commented on the manuscript; L.B.F.M.W. was heavily involved in designing the study and provided much input into the scientific discussion; B.V. was involved in the data reduction and commented on the manuscript; M.M. was involved in the modelling of the temperature structure and opacities and commented on the manuscript; G.L.P. was responsible for obtaining the observations and commented on the manuscript; C.D. was essential in the scientific discussions and helped in obtaining the observations; L.D. was essential in the scientific

discussion and in writing the manuscript. The rest of the authors all contributed equally to obtaining the observations, having scientific discussions and in commenting on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.L.d.V. (bldevries.science@gmail.com).

Stabilizing Rabi oscillations in a superconducting qubit using quantum feedback

R. Vijay¹, C. Macklin¹, D. H. Slichter^{1†}, S. J. Weber¹, K. W. Murch¹, R. Naik¹, A. N. Korotkov² & I. Siddiqi¹

The act of measurement bridges the quantum and classical worlds by projecting a superposition of possible states into a single (probabilistic) outcome. The timescale of this ‘instantaneous’ process can be stretched using weak measurements^{1,2}, such that it takes the form of a gradual random walk towards a final state. Remarkably, the interim measurement record is sufficient to continuously track and steer the quantum state using feedback^{3–8}. Here we implement quantum feedback control in a solid-state system, namely a superconducting quantum bit (qubit) coupled to a microwave cavity⁹. A weak measurement of the qubit is implemented by probing the cavity with microwave photons, maintaining its average occupation at less than one photon. These photons are then directed to a high-bandwidth, quantum-noise-limited amplifier^{10,11}, which allows real-time monitoring of the state of the cavity (and, hence, that of the qubit) with high fidelity. We demonstrate quantum feedback control by inhibiting the decay of Rabi oscillations, allowing them to persist indefinitely¹². Such an ability permits the active suppression of decoherence and enables a method of quantum error correction based on weak continuous measurements^{13,14}. Other applications include quantum state stabilization^{4,7,15}, entanglement generation using measurement¹⁶, state purification¹⁷ and adaptive measurements^{18,19}.

Feedback protocols in classical systems, from antilock brakes to pacemakers, use the outcome of a measurement to stabilize the system about a desired state. The operation of such feedback protocols is predicated on the idea that measurement does not alter the state of the system. This is no longer true in quantum mechanics, where measurement is necessarily invasive¹. In the Copenhagen interpretation, a quantum object can exist simultaneously in more than one eigenstate of the measurement operator until observed—Schrödinger’s celebrated ‘dead-and-alive’ cat being the quintessential hypothetical example²⁰. The reality of the situation is established by the act of measurement, which forces the system ‘instantaneously’ into one of these eigenstates in a probabilistic fashion (the ‘measurement back-action’). Therefore, this back-action must be accounted for when developing a feedback protocol to stabilize a quantum system, such as a qubit.

One solution is to use weak measurements^{1,2}, where the rate (Γ_{meas}) at which information is extracted is deliberately limited, thereby slowing down the qubit’s random walk towards an eigenstate. Integral to this scheme is a detector with efficiency $\eta_{\text{det}} = \Gamma_{\text{meas}}/\Gamma_{\phi} \approx 1$, where Γ_{ϕ} is the ensemble-averaged dephasing rate due to measurement back-action²¹. The high detector efficiency allows us to track the qubit continuously, and steer it to a desired state using real-time feedback.

The experimental set-up is shown in Fig. 1. Our quantum system (Fig. 1b) is an anharmonic oscillator realized by a capacitively shunted Josephson junction, dispersively coupled to a three-dimensional microwave cavity²². We use its two lowest energy levels to form a qubit (transmon²³) with a transition frequency of $\omega_{01}/2\pi = 5.4853$ GHz. The cavity resonant frequency with the qubit in the ground state is $\omega_c/2\pi = 7.2756$ GHz. The strongly coupled output port sets the cavity linewidth $\kappa/2\pi = 13.4$ MHz, and control and measurement signals are injected via the weakly coupled input port (Fig. 1a, b). The qubit–cavity

coupling results in a state-dependent phase shift ($\Delta\phi = 2\tan^{-1}(2\chi/\kappa) = 12^\circ$, $\chi/2\pi = 0.687$ MHz) of the cavity output field^{9,24}, with the state information contained in one quadrature of the signal. The cavity output is sent to a near-noiseless ($\eta_{\text{det}} \approx 1$) phase-sensitive parametric

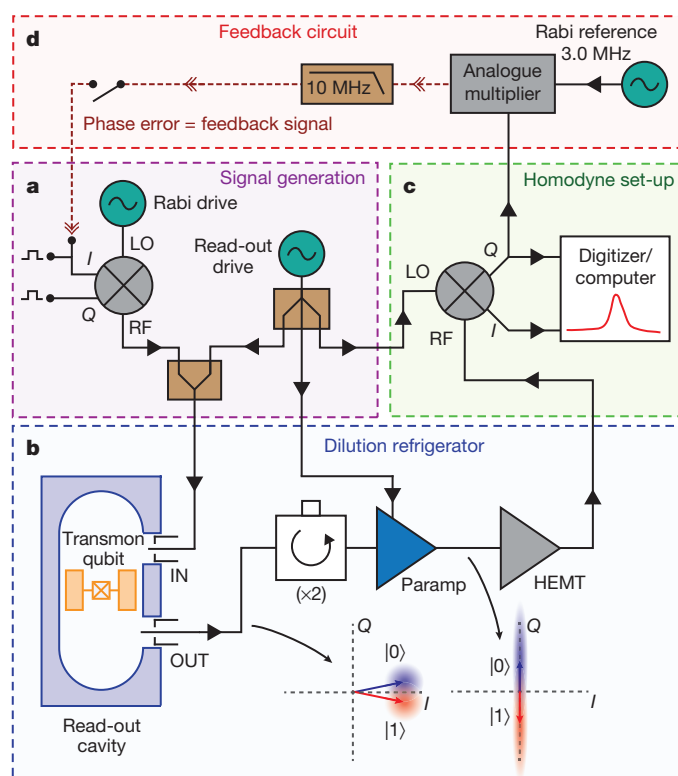


Figure 1 | Experimental set-up. **a**, Signal generation set-up. One generator provides the Rabi drive at the a.c. Stark-shifted qubit frequency ($\omega_{01} - 2\chi\hbar$), and the output of another generator at 7.2749 GHz is split to create the measurement signal, param drive and local oscillator. The relative amplitudes and phases of these three signals are controlled by variable attenuators and phase shifters (not shown). *I*, in-phase component; *Q*, quadrature component; LO, local oscillator; RF, radio frequency. **b**, Simplified version of the cryogenic part of the experiment; all components are at 30 mK (except for the high-electron-mobility transistor (HEMT) amplifier, which is at 4 K). The combined qubit and measurement signals enter the weakly coupled cavity port, interact with the qubit and leave from the strongly coupled port. The output passes through two isolators (which protect the qubit from the strong param drive), is amplified and then continues to the demodulation set-up. The coherent state at the output of the cavity for the ground and excited states is shown schematically before and after parametric amplification. **c**, **d**, The amplified signal is homodyne-detected and the two quadratures are digitized (**c**). The amplified quadrature (*Q*) is split off and sent to the feedback circuit (**d**), where it is multiplied with the Rabi reference signal. The product is low-pass-filtered and fed back to the *IQ* mixer in **a** to modulate the Rabi drive amplitude.

¹Quantum Nanoelectronics Laboratory, Department of Physics, University of California, Berkeley, California 94720, USA. ²Department of Electrical Engineering, University of California, Riverside, California 92521, USA. [†]Present address: Time and Frequency Division, National Institute of Standards and Technology, Boulder, Colorado 80305, USA.

amplifier^{10,11} (paramp), which boosts the relevant quadrature to a level compatible with classical circuitry. The paramp output is further amplified and homodyne-detected (Fig. 1c) such that the amplified quadrature (Q) contains the final measurement signal.

We obtain Rabi oscillations with the cavity continuously excited at $\omega_r/2\pi = 7.2749$ GHz ($\omega_r \approx \omega_c - \chi$) with a mean cavity photon occupation (\bar{n}) that controls the measurement strength (see Supplementary Information, section II, for calibration of \bar{n}). The Rabi drive at the a.c. Stark-shifted²⁵ qubit frequency ($\omega_{01} - 2\chi\bar{n}$) is turned on for a fixed duration, τ_m . The amplitude is adjusted to yield a Rabi frequency of $\Omega_R/2\pi = 3$ MHz. First we average 10^4 measurement traces to obtain a conventional ensemble-averaged Rabi oscillation trace (Fig. 2a). Even though the qubit is continuously oscillating between its ground and excited states, the oscillation phase diffuses, primarily owing to measurement back-action. As a result, the averaged oscillation amplitude decays over time, but the frequency domain response retains a signature of these oscillations²⁶. We Fourier-transform the individual measurement traces and plot the averaged spectrum (Fig. 2b, blue trace). A peak, centred at 3 MHz and with a full-width at half-maximum of $\Gamma/2\pi$, is observed and remains unchanged even when τ_m is much longer than the decay time of the ensemble-averaged oscillations. A plot of $\Gamma/2\pi$ for different measurement strengths (in units of \bar{n}) is shown in Fig. 2c. As expected in the dispersive regime, Γ and \bar{n} are linearly related²⁵. The vertical offset is dominated by pure environmental dephasing, $\Gamma_{\text{env}}/2\pi$, but has contributions from qubit relaxation (T_1) and thermal excitation into higher qubit levels; more details can be found in Supplementary Information, sections II and IV(C).

The ratio of the height of the Rabi spectral peak to the height of the noise floor has a theoretical maximum value of four²⁷, corresponding to an ideal measurement with overall efficiency $\eta = 1$. For our set-up, this efficiency can be separated into two contributions as $\eta = \eta_{\text{det}}\eta_{\text{env}}$. The detector efficiency is given by $\eta_{\text{det}} = (1 + 2n_{\text{add}})^{-1}$, with n_{add} being the number of noise photons added by the amplification chain. The

added noise is referenced to the output of the cavity and includes the effect of signal attenuation between the cavity and the paramp. The effect of environmental dephasing, Γ_{env} , is modelled using $\eta_{\text{env}} = (1 + \Gamma_{\text{env}}/\Gamma_\phi)^{-1}$. The best measurement efficiency we obtain experimentally is $\eta = 0.40$, with $\eta_{\text{det}} = 0.46$ and $\eta_{\text{env}} = 0.87$; more details can be found in Supplementary Information, section III.

We now discuss the quantum feedback protocol, which is motivated by the classical phase-locked loop used for stabilizing an oscillator. The amplified quadrature is multiplied by a Rabi reference signal with frequency $\Omega_0/2\pi = 3$ MHz using an analogue multiplier (Fig. 1d). The output of this multiplier is low-pass-filtered and yields a signal proportional to the sine of the phase difference, θ_{err} , between the 3-MHz reference and the 3-MHz component of the amplified quadrature. This ‘phase error’ signal is fed back to control the Rabi frequency Ω_R by modulating the Rabi drive strength with an upconverting IQ mixer (Fig. 1a). The amplitude of the reference signal controls the dimensionless feedback gain, F , through the expression $\Omega_{\text{fb}}/\Omega_R = -F\sin(\theta_{\text{err}})$, where Ω_{fb} is the change in Rabi frequency due to feedback. Figure 2d shows the ensemble-averaged, feedback-stabilized oscillation, which persists for much longer than the original oscillation in Fig. 2a. In fact, within the limits imposed by our maximum data acquisition time of 20 ms, these oscillations persist indefinitely. The red trace in Fig. 2b shows the corresponding averaged spectra. The needle-like peak at 3 MHz is the signature of the stabilized Rabi oscillations.

To confirm the quantum nature of the feedback-stabilized oscillations, we perform state tomography on the qubit²⁸. We stabilize the dynamical qubit state, stop the feedback and Rabi driving after a fixed time ($80 \mu\text{s} + \tau_{\text{tomo}}$ after starting the Rabi drive), and then measure the projection of the quantum state along one of three orthogonal axes. This is done using strong measurements (by increasing \bar{n}) with high single-shot fidelity¹¹. This allows us to remove any data points where the qubit was found in the second excited state (Supplementary Information, section IV(C)). By repeating this many times, we can determine $\langle\sigma_X\rangle$, $\langle\sigma_Y\rangle$ and $\langle\sigma_Z\rangle$,

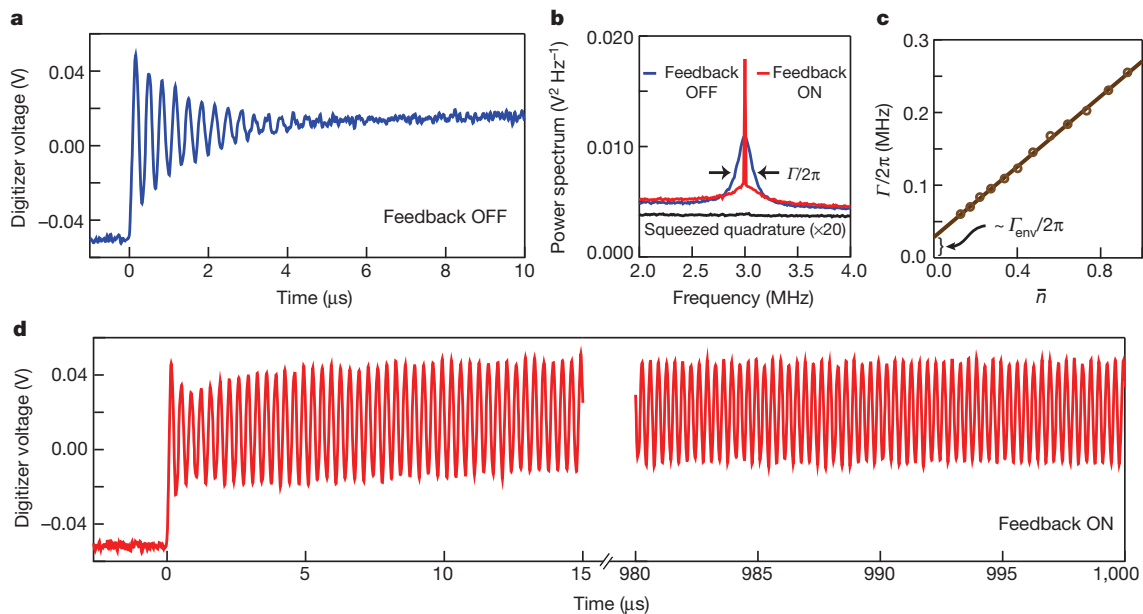


Figure 2 | Rabi oscillations and feedback. **a**, We average 10^4 measurement traces using weak continuous measurement and simultaneous Rabi driving to obtain ensemble-averaged Rabi oscillations that decay in time as a result of ensemble dephasing. **b**, Averaged Fourier transforms of the individual measurement traces from **a**. The spectrum shows a peak at the Rabi frequency (blue trace) with a full-width at half-maximum of $\Gamma/2\pi$. The grey trace shows an identically prepared spectrum for the squeezed quadrature (multiplied by 20 for clarity), which contains no qubit state information. **c**, $\Gamma/2\pi$ plotted as a function of cavity photon occupation, \bar{n} (measurement strength), showing the

expected linear dependence. The vertical offset is dominated by pure environmental dephasing, $\Gamma_{\text{env}}/2\pi$, but has contributions from qubit relaxation (T_1) and thermal excitation into higher qubit levels. **d**, Feedback-stabilized, ensemble-averaged Rabi oscillations, which persist for much longer times than those without feedback (**a**). The corresponding spectrum, shown in **b**, has a needle-like peak at the Rabi reference frequency (red trace). The slowly changing mean level in the Rabi oscillation traces in **a** and **d** is due to the thermal transfer of population into the second excited state of the qubit. See Supplementary Information, section IV(C), for more details.

the three components of the Bloch vector for the ensemble qubit state. Figure 3a shows a plot of the Bloch vector components for different time points (τ_{tomo}) over one oscillation period ($2\pi\Omega_0$). The Y and Z components are well fitted by a sinusoidal function, whereas the X component is nearly zero as expected for a coherent Rabi oscillation about the X axis. The imperfect efficiency of the feedback process is reflected in the non-unit amplitude of these oscillations. This feedback efficiency, D , is given by the time-averaged scalar product of the desired and actual state vectors on the Bloch sphere (Supplementary Information, section IV(A)). In our experiment, the measurement is weak enough that the stabilized oscillations are sinusoidal and D is approximately equal to the amplitude of these oscillations.

In Fig. 3b, we plot D (red squares) versus the dimensionless feedback gain, F . We find a maximum value of $D = 0.45$ for the optimal choice of F . Ideally, feedback efficiency improves with measurement strength ($\eta_{\text{env}} \rightarrow 1$) but requires correspondingly larger feedback loop bandwidth. Thus, in the presence of finite feedback bandwidth and loop delay, there exists an optimal measurement strength, which for our experiment was $\Gamma_\phi/2\pi = 0.134$ MHz. The dashed black line in Fig. 3b is a plot of the theoretical expression for D , given by

$$D = 2 \left(\frac{1}{\eta} \frac{F}{\Gamma/\Omega_0} + \frac{\Gamma/\Omega_0}{F} \right)^{-1} \quad (1)$$

and is derived using a simple analytical theory based on the Bayesian formalism for the qubit state trajectory (Supplementary Information, section IV(A)). This expression does not account for finite feedback bandwidth, loop delays in the circuit or qubit relaxation. The maximum value, $D_{\text{max}} = \sqrt{\eta}$, is obtained for an optimal feedback gain of $F_{\text{opt}} = \sqrt{\eta} \Gamma/\Omega_0$. A value of $D_{\text{max}} < 1$ implies that the stabilized state is a mixed state; this occurs for $\eta < 1$, implying that we have incomplete information about the qubit state. To account for the finite loop delay (250 ns), feedback bandwidth (10 MHz) and qubit relaxation ($T_1 = 20 \mu\text{s}$), we performed full numerical simulations of the Bayesian equations for qubit evolution (Supplementary Information, section IV(B)). The results are shown as a black solid line in Fig. 3b and agree well with our experimental data.

In our experiment, even though the system being controlled is quantum and subject to measurement back-action, we essentially treat it as a classical oscillator and successfully apply a feedback protocol based completely on classical intuition. This can be done because the feedback signal achieves near-perfect cancellation of the random measurement back-action for optimal F . Although it is not true in general, in this particular scheme improvements in the feedback efficiency from full reconstruction of the quantum state⁸ are small. Furthermore, it is possible to approach a pure state with $D = 1$ by ensuring that $\eta = 1$ and eliminating feedback loop delay.

We have demonstrated a continuous analogue feedback scheme to stabilize Rabi oscillations in a superconducting qubit, allowing them to persist indefinitely. The efficiency of the feedback is limited primarily by signal attenuation and loop delay, and could be improved in the near future with the development of on-chip paramps and cryogenic electronics to lessen the effects of attenuation and delay, respectively. We anticipate that our present technology can be extended to entangled qubits to provide another route to quantum error correction based on weak continuous measurements^{13,14}. Such methods might be advantageous in architectures where strong measurements can cause qubit state mixing²⁹. This development may be the start of a new era of measurement-based quantum control for solid-state quantum information processing^{4,7,15–19}.

METHODS SUMMARY

The transmon qubit was fabricated on a bare, high-resistivity Si wafer using electron-beam lithography and double-angle aluminium evaporation with an intervening oxidation step. The qubit is a single Josephson junction connecting two rectangular paddles ($420 \mu\text{m} \times 600 \mu\text{m}$) that provide the shunting capacitance and coupling to the cavity. The cavity was machined out of 6061 aluminium alloy. The quality factor of the cavity was adjusted by controlling the length of the centre conductor of the SMA coaxial connector protruding into the cavity volume. These lengths were chosen to give strong coupling at the output port and weak coupling at the input port, resulting in a net power transmission on resonance of -20 to -30 dB. Qubit rotations around the X and Y axes of the Bloch sphere for state tomography were performed using resonant microwave pulses. The strong measurement used in state tomography was implemented with a 800-ns read-out pulse with an amplitude corresponding to a mean cavity occupation of $\bar{n} \approx 11$.

Received 18 May; accepted 17 August 2012.

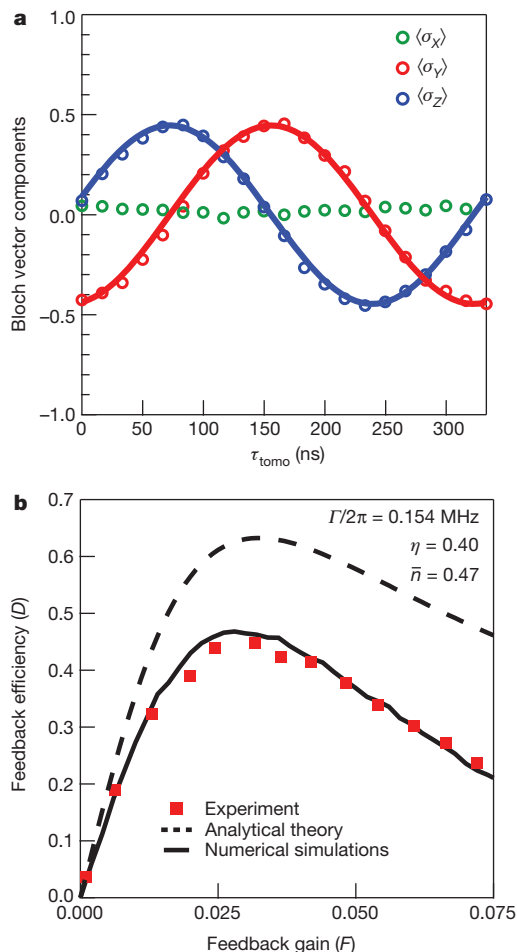


Figure 3 | Tomography and feedback efficiency. **a**, Quantum state tomography of the feedback-stabilized state. We plot $\langle \sigma_x \rangle$, $\langle \sigma_y \rangle$ and $\langle \sigma_z \rangle$ for different time points τ_{tomo} in one full Rabi oscillation of the qubit. The solid lines are sinusoidal fits. The magnitude of these sinusoidal oscillations is approximately equal to the feedback efficiency, $D = 0.45$. **b**, D plotted as a function of the dimensionless feedback gain, F . Solid red squares show experimental data with a maximum value of $D = 0.45$. The dashed black line is a plot of equation (1) with $\eta = 0.40$ and $\Gamma/2\pi = 0.154$ MHz ($\bar{n} = 0.47$, $\Gamma_{\text{env}}/2\pi = 0.020$ MHz), whereas the solid black line is obtained from full numerical simulations of the Bayesian equations including finite loop delay (250 ns) and feedback bandwidth (10 MHz).

- Wiseman, H. M. & Milburn, G. J. *Quantum Measurement and Control* (Cambridge Univ. Press, 2009).
- Gardiner, C. W. & Zoller, P. *Quantum Noise* (Springer, 2004).
- Wiseman, H. M. & Milburn, G. J. Quantum theory of optical feedback via homodyne detection. *Phys. Rev. Lett.* **70**, 548–551 (1993).
- Hofmann, H. F., Mahler, G. & Hess, O. Quantum control of atomic systems by homodyne detection and feedback. *Phys. Rev. A* **57**, 4877–4888 (1998).
- Korotkov, A. N. Selective quantum evolution of a qubit state due to continuous measurement. *Phys. Rev. B* **63**, 115403 (2001).
- Smith, W. P., Reiner, J. E., Orozco, L. A., Kuhr, S. & Wiseman, H. M. Capture and release of a conditional state of a cavity QED system by quantum feedback. *Phys. Rev. Lett.* **89**, 133601 (2002).
- Gillett, G. G. *et al.* Experimental feedback control of quantum systems using weak measurements. *Phys. Rev. Lett.* **104**, 080503 (2010).

8. Sayrin, C. *et al.* Real-time quantum feedback prepares and stabilizes photon number states. *Nature* **477**, 73–77 (2011).
9. Blais, A., Huang, R.-S., Wallraff, A., Girvin, S. M. & Schoelkopf, R. J. Cavity quantum electrodynamics for superconducting electrical circuits: an architecture for quantum computation. *Phys. Rev. A* **69**, 062320 (2004).
10. Hatridge, M., Vijay, R., Slichter, D. H., Clarke, J. & Siddiqi, I. Dispersive magnetometry with a quantum limited SQUID parametric amplifier. *Phys. Rev. B* **83**, 134501 (2011).
11. Vijay, R., Slichter, D. H. & Siddiqi, I. Observation of quantum jumps in a superconducting artificial atom. *Phys. Rev. Lett.* **106**, 110502 (2011).
12. Ruskov, R. & Korotkov, A. N. Quantum feedback control of a solid-state qubit. *Phys. Rev. B* **66**, 041401 (2002).
13. Ahn, C., Doherty, A. C. & Landahl, A. J. Continuous quantum error correction via quantum feedback control. *Phys. Rev. A* **65**, 042301, 2002..
14. Tomberg, L. & Johansson, G. High-fidelity feedback-assisted parity measurement in circuit QED. *Phys. Rev. A* **82**, 012329 (2010).
15. Wang, J. & Wiseman, H. M. Feedback-stabilization of an arbitrary pure state of a two-level atom. *Phys. Rev. A* **64**, 063810 (2001).
16. Ruskov, R. & Korotkov, A. N. Entanglement of solid-state qubits by measurement. *Phys. Rev. B* **67**, 241305 (2003).
17. Combes, J. & Jacobs, K. Rapid state reduction of quantum systems using feedback control. *Phys. Rev. Lett.* **96**, 010504 (2006).
18. Jacobs, K. Feedback control for communication with non-orthogonal states. *Quantum Inf. Comput.* **7**, 127–138 (2007).
19. Cook, R. L., Martin, P. J. & Geremia, J. M. Optical coherent state discrimination using a closed-loop quantum measurement. *Nature* **446**, 774–777 (2007).
20. Schrödinger, E. The present situation in quantum mechanics. *Proc. Am. Phil. Soc.* **124**, 323–338 (1980).
21. Clerk, A. A., Devoret, M. H., Girvin, S. M., Marquardt, F. & Schoelkopf, R. J. Introduction to quantum noise, measurement, and amplification. *Rev. Mod. Phys.* **82**, 1155–1208 (2010).
22. Paik, H. *et al.* Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture. *Phys. Rev. Lett.* **107**, 240501 (2011).
23. Koch, J. *et al.* Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).
24. Wallraff, A. *et al.* Approaching unit visibility for control of a superconducting qubit with dispersive readout. *Phys. Rev. Lett.* **95**, 060501 (2005).
25. Schuster, D. I. *et al.* ac Stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field. *Phys. Rev. Lett.* **94**, 123602 (2005).
26. Palacios-Laloy, A. *et al.* Experimental violation of a Bell's inequality in time with weak measurement. *Nature Phys.* **6**, 442–447 (2010).
27. Korotkov, A. N. & Averin, D. V. Continuous weak measurement of quantum coherent oscillations. *Phys. Rev. B* **64**, 165310 (2001).
28. Steffen, M. *et al.* State tomography of capacitively shunted phase qubits with high fidelity. *Phys. Rev. Lett.* **97**, 050502 (2006).
29. Slichter, D. H. *et al.* Measurement-induced qubit state mixing in circuit QED from upconverted dephasing noise. Preprint at <http://arxiv.org/abs/1206.6946> (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Sarovar for several discussions and Z. Mineev for assistance with numerical simulations. This research was supported in part (R.V., C.M. and I.S.) by the US Army Research Office (W911NF-11-1-0029) and the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Office (K.W.M., S.J.W. and A.N.K.). All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI or the US government. D.H.S. acknowledges support from a Hertz Foundation Fellowship endowed by Big George Ventures. A.N.K. also acknowledges funding from an ARO MURI.

Author Contributions R.V., C.M. and D.H.S. performed the experiment, which is based on a proposal by A.N.K. R.V. analysed the data, performed numerical simulations and wrote the manuscript. S.J.W. and K.W.M. fabricated the qubit and cavity. R.N. helped with cavity design by performing electromagnetic simulations. A.N.K. provided theoretical support and helped with numerical simulations. All authors helped in editing the manuscript. All work was carried out under the supervision of I.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.V. (rvijay@berkeley.edu) or I.S. (irfan@berkeley.edu).

Sulphate–climate coupling over the past 300,000 years in inland Antarctica

Yoshinori Iizuka¹, Ryu Uemura², Hideaki Motoyama³, Toshitaka Suzuki⁴, Takayuki Miyake^{3†}, Motohiro Hirabayashi³ & Takeo Hondoh¹

Sulphate aerosols, particularly micrometre-sized particles of sulphate salt and sulphate-adhered dust, can act as cloud condensation nuclei, leading to increased solar scattering that cools Earth's climate^{1,2}. Evidence for such a coupling may lie in the sulphate record from polar ice cores, but previous analyses of melted ice-core samples have provided only sulphate ion concentrations, which may be due to sulphuric acid³. Here we present profiles of sulphate salt and sulphate-adhered dust fluxes over the past 300,000 years from the Dome Fuji ice core in inland Antarctica. Our results show a nearly constant flux of sulphate-adhered dust through glacial and interglacial periods despite the large increases in total dust flux during glacial maxima⁴. The sulphate salt flux, however, correlates inversely with temperature, suggesting a climatic coupling between particulate sulphur and temperature. For example, the total sulphate salt flux during the Last Glacial Maximum averages $5.78 \text{ mg m}^{-2} \text{ yr}^{-1}$, which is almost twice the Holocene value. Although it is based on a modern analogue with considerable uncertainties when applied to the ice-core record, this analysis indicates that the glacial-to-interglacial decrease in sulphate would lessen the aerosol indirect effects on cloud lifetime and albedo, leading to an Antarctic warming of 0.1 to 5 kelvin.

The climate of the past 430 kyr is guided by large-amplitude 100-kyr glacial–interglacial cycles⁵. Such cycles are triggered by the Earth's orbital changes and then amplified by various land–atmosphere interactions¹. Of these, the influence of aerosol radiative forcing¹ is particularly hard to quantify accurately. The low aerosol concentrations of the southern high-latitude region make the direct effect of aerosol radiative forcing negligible, but the indirect effect, in which the aerosol acts as cloud condensation nuclei (CCN), could be meaningful^{6,7}. Many CCN consist of sulphate salt, either by itself or mixed with silicate materials⁶ (Supplementary Fig. 1), which after precipitating to the ground can be preserved in the ice.

Aerosols in the Antarctic region contain little anthropogenic and ammonium sulphates⁸, instead being dominated by sulphuric acid (H_2SO_4), sodium sulphate (Na_2SO_4) and calcium sulphate⁹ (CaSO_4). The Na_2SO_4 arises mostly from the reaction of sodium chloride (NaCl) with H_2SO_4 that comes mainly from marine biological activity¹⁰. The CaSO_4 arises from terrestrial gypsum and also from a reaction in aerosol between H_2SO_4 and calcium carbonate¹⁰. These salts primarily form during their transport through the atmosphere¹¹. Once in the ice, sulphate salts are good proxies for past atmospheric chemistry because, unlike the more mobile, volatile compounds such as nitric acid¹² (HNO_3) and liquid H_2SO_4 (ref. 8), they are largely unaffected by post-depositional processes.

Analyses^{3,13} of the sulphate ion (SO_4^{2-}) record in Antarctic ice cores show that almost all non-seasalt sulphate ions in inland Antarctic ice cores come from marine biological activity¹³. However, because the ice samples were melted before measurement, those studies could not definitively separate the H_2SO_4 flux from the sulphate salt (Na_2SO_4

and CaSO_4) flux. Even less is known about how much of the sulphate salt mixed with silicate mineral dust. Here we present a complete record of the sulphate salt and sulphate-adhered dust fluxes into inland Antarctica over the past 300 kyr.

To determine the sulphate salt and seasalt fluxes (by sea salt we mean only NaCl), we combined two methods: an ion analysis method⁹, here applied to the 300-kyr record in the Dome Fuji ice core, and an elemental and compositional analysis of single particles extracted from the same ice record, using a sublimation method^{11,14}.

We found that particles (Supplementary Fig. 2) in glacial inception and interglacial periods can be distinguished from those in glacial maxima by their $\text{Na}_2\text{SO}_4/\text{CaSO}_4$ and $\text{NaCl}/\text{Na}_2\text{SO}_4$ mass ratios. During glacial maxima, $\text{Na}_2\text{SO}_4/\text{CaSO}_4$ values are particularly low, whereas $\text{NaCl}/\text{Na}_2\text{SO}_4$ values are high (Fig. 1). With a squared correlation coefficient of $r^2 = 0.76$ and a regression slope close to 1 (0.85 ± 0.17 in Supplementary Fig. 3), the $\text{Na}_2\text{SO}_4/\text{CaSO}_4$ mass ratio correlates well with that⁹ from the ion chromatograph analyses.

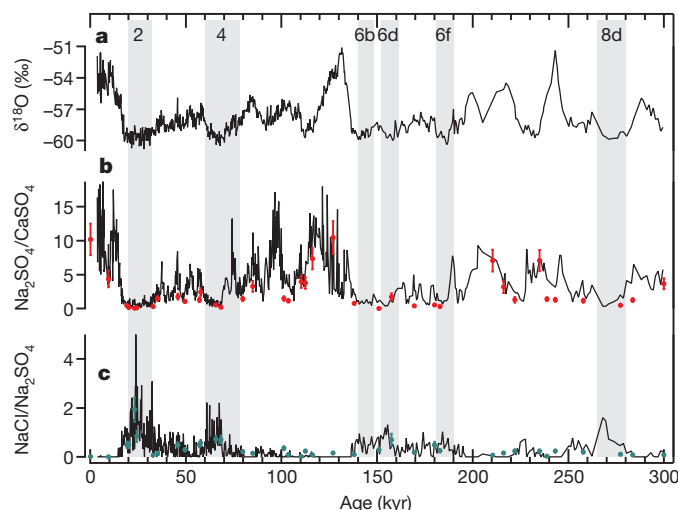


Figure 1 | Oxygen isotope and salt mass ratios of the past 300 kyr (before AD 2000) in the Dome Fuji ice core. **a**, Oxygen isotope ratio of ice (expressed as $\delta^{18}\text{O} = (^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (^{18}\text{O}/^{16}\text{O})_{\text{standard}} - 1$, where the standard is Vienna Standard Mean Ocean Water) from ref. 27 ($n = 691$; uncertainty, $<0.1\text{‰}$). Grey bars indicate glacial maxima (Marine Isotope Stages 2, 4, 6b, 6d, 6f and 8d). **b**, $\text{Na}_2\text{SO}_4/\text{CaSO}_4$ salt mass ratio from the energy-dispersive spectroscopy (EDS) method (single-particle analysis) applied to 38 filter samples (solid circles) and measured from ion concentrations²⁷ of 691 melt samples (line). Vertical error bars are uncertainties (coefficient of variation) in the EDS ratios. Numerical values are in Supplementary Information. The uncertainty in the ion-deduced ratios ranges from 36 to 49% for average glacial and interglacial conditions. **c**, Same as **b** except data shows the $\text{NaCl}/\text{Na}_2\text{SO}_4$ salt mass ratio. The uncertainty in the ion-deduced ratios ranges from 34 to 19% for average glacial and interglacial conditions.

¹Institute of Low Temperature Science, Hokkaido University, Sapporo 060-0819, Japan. ²Department of Chemistry, Biology and Marine Science, Faculty of Science, University of the Ryukyus, Okinawa 903-0213, Japan. ³National Institute of Polar Research, Tokyo 190-8518, Japan. ⁴Department of Earth and Environmental Sciences, Faculty of Science, Yamagata University, Yamagata 990-8560, Japan. [†]Present address: School of Environmental Science, The University of Shiga Prefecture, Shiga 522-8533, Japan.

Because the ion-deduced CaSO_4 flux is well confirmed from the relation between Ca^{2+} and particle components^{14,15} in the Dome Fuji ice core, we can confidently use the ion-deduced Na_2SO_4 and CaSO_4 fluxes. The regression slope of $\text{NaCl}/\text{Na}_2\text{SO}_4$ equals 0.84 (Supplementary Fig. 3), which, considering the $\text{Na}_2\text{SO}_4/\text{CaSO}_4$ slope of 0.85, indicates that the ion chromatograph method underestimated the NaCl flux by up to 29% ($1 - 0.85 \times 0.84$). This underestimate may be due to a poor understanding of mixed nitrate and chloride salt formation⁹. Nevertheless, the high correlation ($r^2 = 0.73$) suggests that the NaCl variations during the 300-kyr time series can be reconstructed well.

The resulting ion-chromatograph-derived CaSO_4 and NaCl fluxes to the ground surface vary greatly over the past 300 kyr, both showing high values during glacial maxima (Fig. 2). The high CaSO_4 flux agrees with the high dust flux from South America during these stages^{4,16}, which suggests that the higher dust flux increases Ca^{2+} sulphatization and thus consumes more H_2SO_4 in the atmosphere¹⁷.

During interglacials, both the NaCl flux and the Na_2SO_4 flux are low, probably because the reaction between NaCl and H_2SO_4 to produce Na_2SO_4 is limited by the low NaCl flux^{8,17}. Moreover, the NaCl/ Na_2SO_4 ratio is low during interglacials and glacial inceptions but high in glacial maxima (Fig. 1), a trend predicted by the reaction between NaCl and H_2SO_4 (ref. 17). A low ratio in interglacials and glacial inceptions indicates excess H_2SO_4 in the Antarctic atmosphere (Fig. 2), causing the sulphate salt flux to be controlled mainly by seasalt flux^{8,9}. However, a high ratio in glacial maxima indicates excess seasalt flux⁹, causing the sulphate salt flux to be controlled mainly by H_2SO_4 flux. Consistent with this interpretation, we measured very little H_2SO_4 during glacial maxima (Fig. 2), as nearly all SO_4^{2-} was associated with sulphate salt. In contrast, during interglacial periods and

glacial inceptions, the H_2SO_4 contribution is relatively high, meaning that the SO_4^{2-} flux is not a good measure of the sulphate salt flux during these periods.

We also estimated the flux of sulphate-adhered dust, that is, particles with both silicon and sulphate. Such particles probably act as CCN^{6,18} and thus influence climate. We refer to the collection of all particles containing silicon as the 'total dust', which is often measured in ice-core studies^{4,16}. Because the total dust flux correlates to the non-seasalt calcium ion flux^{3,4}, we examined the time variation of the ratio of sulphate-adhered dust to total dust in comparison with the ratio of sulphate salt flux to non-seasalt calcium ion flux (Supplementary Fig. 4). These ratios are correlated, both being high during glacial inceptions but low during glacial maxima. We use this correlation together with the measured total-dust flux from the melt samples to obtain the sulphate-adhered dust flux.

The sulphate-adhered dust flux averages $0.10 \text{ mg m}^{-2} \text{ yr}^{-1}$ through glacial and interglacial periods (Fig. 3a). This constant trend occurs despite the total dust flux increasing by a factor of 50 (up to $\sim 5 \text{ mg m}^{-2} \text{ yr}^{-1}$) between interglacials and glacial maxima. Because of the constant trend, no significant correlation occurs between temperature and sulphate-adhered dust flux, suggesting that the sulphate-adhered dust has little influence on the glacial and interglacial temperature changes via the indirect effect of CCN.

In contrast to that of the sulphate-adhered dust, the sulphate salt flux (another measure of CCN) correlates inversely with $\delta^{18}\text{O}$, a temperature proxy (Fig. 3b). This correlation suggests a coupling between temperature and particulate sulphur (through Na_2SO_4 and CaSO_4). The higher sulphate salt flux at lower temperatures also contrasts with the SO_4^{2-} flux, which shows no clear correlation with temperature³.

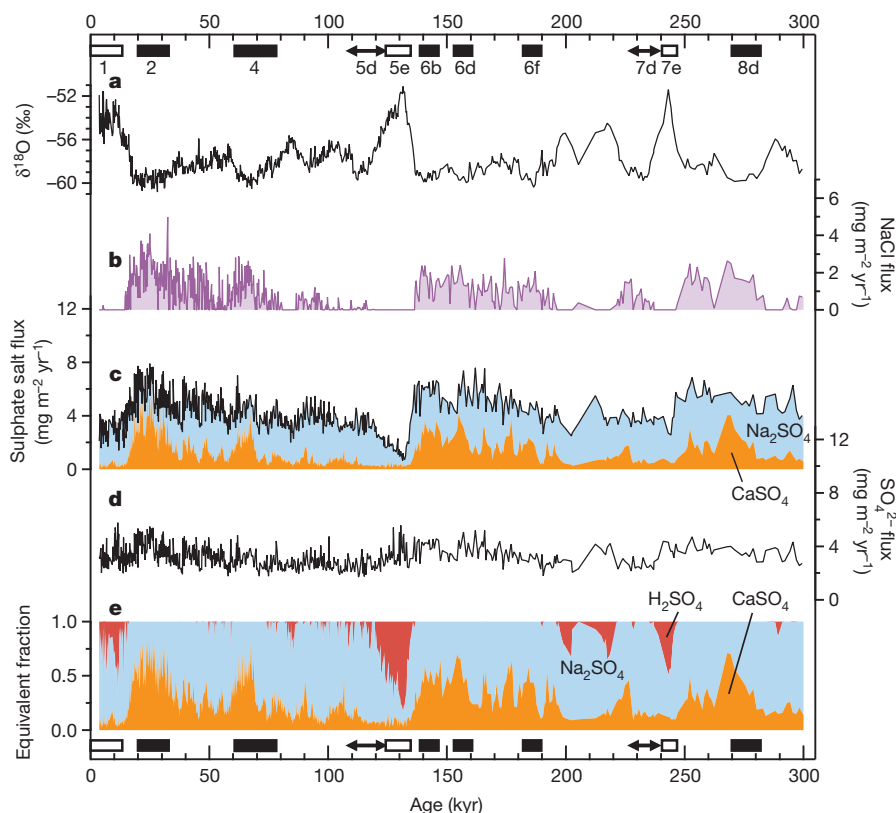


Figure 2 | Sulphate ion and salts fluxes. **a**, Isotopic data as in Fig. 1a. Marine Isotope Stages 1–8d are indicated at top. Open boxes (1, 5e and 7e) are interglacials, arrows (5d and 7d) are glacial inceptions and filled boxes (2, 4, 6b, 6d, 6f and 8d) are glacial maxima. **b**, NaCl flux from ion concentration measurements²⁷ after being verified by the single-particle (EDS) measurements. The uncertainty in the NaCl flux is 17% for average glacial conditions and 10% for interglacial conditions. **c**, Cumulative salt fluxes from ion concentration

measurements²⁷ after being verified by EDS measurements. Na_2SO_4 , blue; CaSO_4 , orange. The Na_2SO_4 flux uncertainty ranges from 17 to 10% for average glacial and interglacial conditions; the corresponding values for CaSO_4 range from 19 to 39%. **d**, The sulphate ion (SO_4^{2-}) flux²⁷. The uncertainty in the SO_4^{2-} flux ranges from 17 to 5.7% for average glacial and interglacial conditions. **e**, Fractional contribution of sulphuric acid (H_2SO_4) and sulphate salt (Na_2SO_4 plus CaSO_4) to the total sulphur flux.

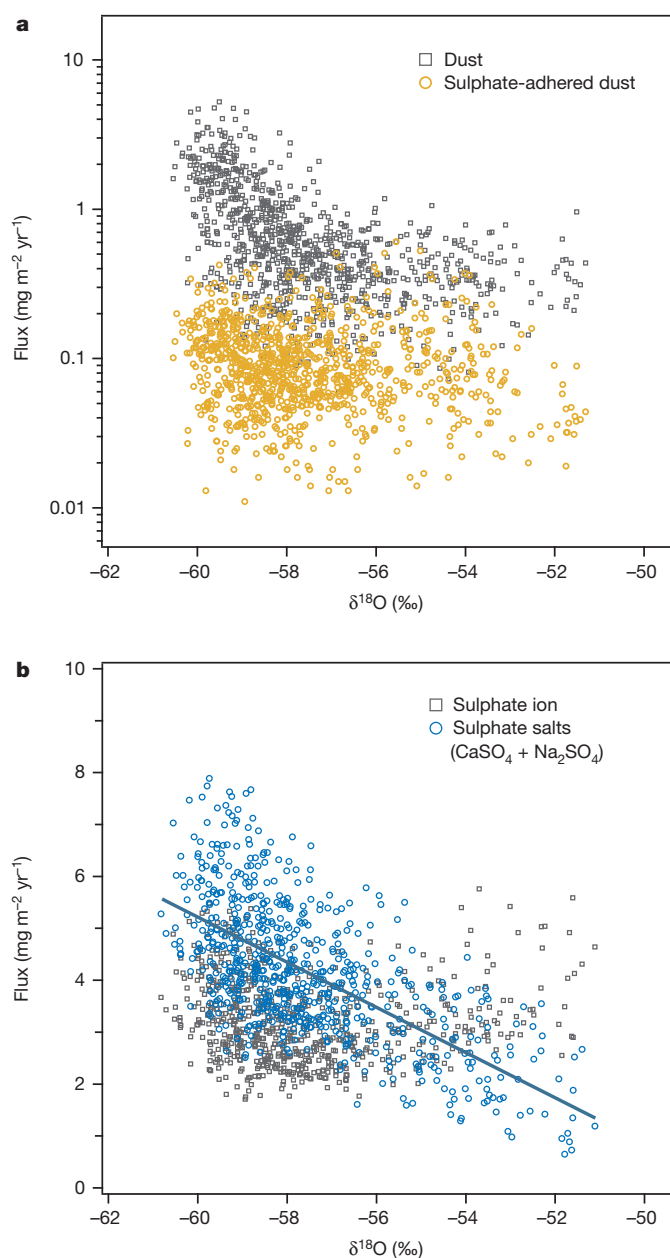


Figure 3 | Correlations of relevant fluxes to the temperature proxy $\delta^{18}\text{O}$.

a, Sulphate-adhered and total dust fluxes²⁸. The uncertainty in the sulphate-adhered dust flux ranges from 31 to 58% for average glacial and interglacial conditions; the corresponding values for the total dust flux are 17 to 14%. **b**, Sulphate ion and sulphate salt fluxes. The linear best-fit line for sulphate salts is $\text{Flux} = -0.44\delta^{18}\text{O} - 20.9$ with $r^2 = 0.42$ ($n = 691$). The correlation is significant ($P < 0.001$). (For sulphate ions, a linear fit gives an r^2 value of only 0.005). The uncertainty in the sulphate ion flux ranges from 17 to 5.7% for average glacial and interglacial conditions; for the sulphate salt flux, the range is 17 to 12%.

For the relatively warm interglacial periods, including the Holocene epoch, the evidence here suggests that Na_2SO_4 is generally controlled by the flux of sea salt from the ocean, which is relatively low. But in glacial maxima, the seasalt and dust fluxes increase, leading to greater particulate sulphur flux. (The flux of sulphate salt exceeds that of SO_4^{2-} because the salt has a higher molar mass.) During such cold periods, the amount of particulate sulphur is limited by the amount of SO_4^{2-} . This SO_4^{2-} comes from marine biogenic sulphur¹³, and marine biological activity thus influences the amount of sulphate salt.

The coupling shown here between temperature and sulphate salt flux provides new evidence against which to test the long-debated¹⁹

hypothesis of Charlson *et al.*²⁰ (CLAW). They proposed that marine biogenic sulphur provides a negative feedback to climate change. Such a feedback requires the sulphur to have a controlling effect on CCN production and requires biogenic sulphur production to respond to climate. Concerning the first requirement, a recent review¹⁹ shows that sulphur does not control CCN in the present interglacial period. Our evidence from the three most recent interglacials agrees with this view because our data indicate that sea salt rather than marine sulphur controls particulate sulphate flux. However, during glacial maxima, the greater sulphate salt flux is determined by the marine sulphur flux, which agrees with the first requirement. However, the CLAW negative feedback requires that the greater sulphate salt flux decreases marine sulphur emissions as a consequence of changes in cloud albedo and surface temperature, which is inconsistent with our finding that SO_4^{2-} flux does not decrease during glacial maxima (Fig. 3). Thus, the second CLAW requirement seems to fail even during glacial maxima.

The correlation between temperature and sulphate salts suggests that the sulphate salts have an indirect aerosol effect on glacial–interglacial temperature changes. To obtain a rough estimate of the radiative forcing caused by the flux change of sulphate salt, we use recent radiative modelling results. To do so, we note that the ratio of sulphate concentration during the Holocene to that during the Last Glacial Maximum roughly equals the ratio of pre-industrial anthropogenic sulphate emissions^{21,22} (Supplementary Fig. 5), and we assume that the sulphate salt concentration changes globally. This assumption is consistent with the finding that the change of dust flux in Antarctica correlates strongly with global dust flux²³.

The resulting model gives net sulphate-induced radiative coolings in the Last Glacial Maximum of -1.85 W m^{-2} (ref. 24) and -2.57 W m^{-2} (ref. 25), producing global climate coolings of 0.1 (ref. 24) and 2.24 K (ref. 25). By applying the polar amplification factor²⁶ (1.3–2.3), the resulting cooling in Antarctica ranges from 0.1 to 5 K as a result only of the indirect aerosol effect. The value 5 K is probably overestimated because other reports² suggest lower radiative cooling. Considering that the temperature change between the Last Glacial Maximum and the Holocene in Antarctica is $\sim 8^\circ\text{C}$ (ref. 26), our estimate highlights the large uncertainty in the climatic impact of indirect aerosol effects from sulphate salts. Thus, the role of sulphate salts and sulphate-adhered dust in producing the indirect effect should be quantified using palaeoclimate models with our new aerosol data and other proxies (for example temperature, CO_2 , CH_4 and total dust) archived in the ice core.

METHODS SUMMARY

To determine the amount of sulphate, chloride and mineral particles in the Dome Fuji ice core, we used a low-temperature sublimation and particle analysis method^{11,14} and compared the results with an analysis⁹ based on ion chromatography data obtained previously²⁷. That analysis determines the salt fluxes of CaSO_4 , Na_2SO_4 plus MgSO_4 , and NaCl plus MgCl_2 from ion concentrations. Results from the two methods agreed well for CaSO_4 and the Na_2SO_4 -plus- MgSO_4 fluxes (Supplementary Fig. 3).

The sublimation method removes volatiles such as H_2O , HCl , HNO_3 and H_2SO_4 (refs 11, 14). In total, we sublimated 38 sampled sections over the 300-kyr period and analysed them using scanning electron microscopy/energy-dispersive spectroscopy. Each sample yielded several hundred particles exceeding $0.4 \mu\text{m}$ in diameter (16,821 particles in total). To determine whether the particles had a mineral, a sulphate or a chloride component, we used a scheme from previous studies^{11,14} that divides non-volatile particles into insoluble and soluble components. We assumed that sodium and sulphur in a given particle came from Na_2SO_4 , whereas sodium and chlorine came from NaCl and calcium and sulphur came from CaSO_4 . Then we calculated the $\text{Na}_2\text{SO}_4/\text{NaCl}$ and $\text{Na}_2\text{SO}_4/\text{CaSO}_4$ mass ratios following the method described in ref. 11. To obtain the number ratio of sulphate-adhered dust to total silicate dust particles, we assumed that a particle with both sulphur and silicon was sulphate-adhered dust and divided the number of such particles by the total number of particles with silicon. Supplementary Information lists the uncertainties.

Received 31 January; accepted 27 June 2012.

- Jansen, E. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 433–498 (Cambridge Univ. Press, 2007).
- Forster, P. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 129–234 (Cambridge Univ. Press, 2007).
- Wolff, E. W. *et al.* Southern Ocean sea-ice extent, productivity and iron flux over the past eight glacial cycles. *Nature* **440**, 491–496 (2006).
- Lambert, F. *et al.* Dust climate couplings over the past 800,000 years from the EPICA Dome C ice core. *Nature* **452**, 616–619 (2008).
- EPICA community members. Eight glacial cycles from an Antarctic ice core. *Nature* **429**, 623–628 (2004).
- Andreae, M. O. *et al.* Internal mixture of sea salt, silicates, and excess sulfate in marine aerosols. *Science* **232**, 1620–1623 (1986).
- Bodhaine, B. A. Aerosol absorption measurements at Barrow, Mauna Loa and the South Pole. *J. Geophys. Res.* **100** (D5), 8967–8975 (1995).
- Iizuka, Y., Hondoh, T. & Fujii, Y. Na₂SO₄ and Mg₂SO₄ salts during the Holocene period derived by high-resolution depth analysis of a Dome Fuji ice core. *J. Glaciol.* **52**, 58–64 (2006).
- Iizuka, Y. *et al.* A relationship between ion balance and the chemical compounds of salt inclusions found in the GRIP and Dome Fuji ice cores. *J. Geophys. Res.* **113**, D07303 (2008).
- Legrand, M. R., Lorius, C., Barkov, N. I. & Petrov, V. N. Vostok (Antarctica) ice core: atmospheric chemistry changes over the last climatic cycle (160,000 years). *Atmos. Environ.* **22**, 317–331 (1988).
- Iizuka, Y. *et al.* The rates of sea salt sulfatization in the atmosphere and surface snow of inland Antarctica. *J. Geophys. Res.* **117**, D04308 (2012).
- Röthlisberger, R. *et al.* Nitrate in Greenland and Antarctic ice cores: a detailed description of post-depositional processes. *Ann. Glaciol.* **35**, 209–216 (2002).
- Kaufmann, P. *et al.* Ammonium and non-sea salt sulfate in the EPICA ice cores as indicator of biological activity in the Southern Ocean. *Quat. Sci. Rev.* **29**, 313–323 (2010).
- Iizuka, Y. *et al.* Constituent elements of insoluble and nonvolatile particles during the Last Glacial Maximum of the Dome Fuji ice core. *J. Glaciol.* **55**, 552–562 (2009).
- Sakurai, T. *et al.* The chemical forms of water-soluble microparticles preserved in the Antarctic ice sheet during Termination I. *J. Glaciol.* **57**, 1027–1032 (2011).
- Delmonte, B. *et al.* Aeolian dust in East Antarctica (EPICA-Dome C and Vostok): provenance during glacial ages over the last 800 kyr. *Geophys. Res. Lett.* **35**, L07703 (2008).
- Röthlisberger, R. *et al.* Limited dechlorination of sea salt aerosols during the last glacial period: evidence from the European Project for Ice Coring in Antarctica (EPICA) Dome C ice core. *J. Geophys. Res.* **108**, 4526 (2003).
- Manktelow, P. T., Carslaw, K. S., Mann, G. W. & Spracklen, D. V. The impact of dust on sulfate aerosol, CN and CCN during an East Asian dust storm. *Atmos. Chem. Phys.* **10**, 365–382 (2010).
- Quinn, P. K. & Bates, T. S. The case against climate regulation via oceanic phytoplankton sulphur emissions. *Nature* **480**, 51–56 (2011).
- Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. Oceanic phytoplankton, atmospheric sulphur, cloud albedo, and climate. *Nature* **326**, 655–661 (1987).
- Penner, J. E. *et al.* in *Climate Change 2001, Working Group I: The Scientific Basis* (eds Houghton, J. T. *et al.*) 289–348 (Cambridge Univ. Press, 2001).
- Fujii, Y. *et al.* 210-year ice core records of dust storms, volcanic eruptions and acidification at Site-J, Greenland. *Mem. Natl Inst. Polar Res. Spec. Issue* **54**, 209–220 (2001).
- Winckler, G. *et al.* Covariant glacial-interglacial dust fluxes in the equatorial Pacific and Antarctica. *Science* **320**, 93–96 (2008).
- Tsai, I. C. *et al.* Sulfur cycle and sulfate radiative forcing simulated from a coupled global climate-chemistry model. *Atmos. Chem. Phys.* **10**, 3693–3709 (2010).
- Lohmann, U. & Feichter, J. Global indirect aerosol effects: a review. *Atmos. Chem. Phys.* **5**, 715–737 (2005).
- Masson-Delmotte, V. *et al.* Past and future polar amplification of climate change: climate model intercomparisons and ice-core constraints. *Clim. Dyn.* **26**, 513–529 (2006).
- Watanabe, O. *et al.* General tendencies of stable isotopes and major chemical constituents of the Dome Fuji deep ice core. *Mem. Natl Inst. Polar Res. Spec. Issue* **57**, 1–24 (2003).
- Fujii, Y., Kohno, M., Matoba, S., Motoyama, H. & Watanabe, O. A 320 k-year record of microparticles in the Dome Fuji, Antarctica ice core measured by laser-light scattering. *Mem. Natl Inst. Polar Res. Spec. Issue* **57**, 46–62 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank all members of the JARE and the Ice Core Consortium for work at Dome Fuji; T. Takemura, R. Kudo, Y. Fujiyoshi and T. Aoki for comments on radiative forcing of aerosols; H. Ohno and T. Sakurai for comments on microparticles in the Dome Fuji ice core; K. Goto-Azuma, T. Kuramoto and H. Yamada for data management of ion and dust concentrations in the Dome Fuji ice core; and Y. Fujii for data on ion concentrations in the Site J firn core, used to calculate sulphate salts. Y.I. thanks M. Furusaki for help with EDS analyses and J. Nelson for help with revising the manuscript. This study was supported by Creative Scientific Research (grant number 14GS0202), Young Scientists (A) (grant number 23680001) and Scientific Research (S) (grant number 21221002), provided by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS) and the Grant for Joint Research Program of the Institute of Low Temperature Science, Hokkaido University.

Author Contributions Y.I. ran the experiments and analysed data on statistical methods. Y.I., R.U., T.S. H.M. and T.H. discussed the results and commented on the manuscript. T.M. and M.H. provided the dust and ion data. Y.I. and R.U. wrote the paper. Y.I., H.M. and T.H. provided financial resources.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.I. (iizuka@lowtem.hokudai.ac.jp).

Natural and anthropogenic variations in methane sources during the past two millennia

C. J. Sapart¹, G. Monteil¹, M. Prokopiou¹, R. S. W. van de Wal¹, J. O. Kaplan², P. Sperlich³, K. M. Krumhardt², C. van der Veen¹, S. Houweling^{1,4}, M. C. Krol¹, T. Blunier³, T. Sowers⁵, P. Martinerie⁶, E. Witrant⁷, D. Dahl-Jensen³ & T. Röckmann¹

Methane is an important greenhouse gas that is emitted from multiple natural and anthropogenic sources. Atmospheric methane concentrations have varied on a number of timescales in the past, but what has caused these variations is not always well understood^{1–8}. The different sources and sinks of methane have specific isotopic signatures, and the isotopic composition of methane can therefore help to identify the environmental drivers of variations in atmospheric methane concentrations⁹. Here we present high-resolution carbon isotope data ($\delta^{13}\text{C}$ content) for methane from two ice cores from Greenland for the past two millennia. We find that the $\delta^{13}\text{C}$ content underwent pronounced centennial-scale variations between 100 BC and AD 1600. With the help of two-box model calculations, we show that the centennial-scale variations in isotope ratios can be attributed to changes in pyrogenic and biogenic sources. We find correlations between these source changes and both natural climate variability—such as the Medieval Climate Anomaly and the Little Ice Age—and changes in human population and land use, such as the decline of the Roman empire and the Han dynasty, and the population expansion during the medieval period.

In the pre-industrial period, methane (CH_4) sources can be divided into three categories on the basis of their stable carbon isotopic signatures: biogenic sources (for example, wetlands, rice paddies and ruminants, mean $\delta^{13}\text{C} = -60 \pm 5\text{‰}$), geological sources¹⁰ (for example, mud volcanoes and microseepages, mean $\delta^{13}\text{C} = -38 \pm 7\text{‰}$) and pyrogenic sources (for example, fires, biofuel and coal burning, mean $\delta^{13}\text{C} = -22 \pm 3\text{‰}$)^{1–3,9}. The isotopic composition of CH_4 in the troposphere is affected by emissions from these sources and by isotope fractionation in the sink mechanisms, primarily OH^\bullet oxidation, with minor contributions from soil removal and stratospheric loss³.

Previous measurements of the $\delta^{13}\text{C}$ value of CH_4 from air trapped in Antarctic ice cores challenged our understanding of the behaviour of pre-industrial CH_4 sources^{1,2}. They indicate that before AD 1500, the contribution of ^{13}C -enriched CH_4 sources (for example, biomass burning) had to be larger than expected for pre-industrial conditions in order to explain the generally high $\delta^{13}\text{C}$ levels during this period^{1–3}. After AD 1500, $\delta^{13}\text{C}$ decreased by 2‰ until AD 1800, followed by an abrupt increase presumably caused by increased fossil fuel emissions associated with the onset of industrialization^{1–3}. Several hypotheses have been proposed to explain the $\delta^{13}\text{C}$ minimum around AD 1800, including a decline in anthropogenic biomass burning in the Americas concurrent with colonial expansion¹, an early rise of ^{13}C -depleted agricultural sources³ or a combination of both².

Our high-resolution $\delta^{13}\text{C}$ data from the NEEM (North Greenland Eemian Ice Drilling programme) and EUROCORE ice cores (see Methods) allow a more detailed reconstruction of global-scale changes in CH_4 sources over the past two millennia (Fig. 1a). Whereas the most distinctive feature of the isotopic record over this period is the minimum around AD 1800 (in agreement with previous studies^{1,2}),

our measurements also reveal three centennial-scale excursions in $\delta^{13}\text{C}$ between 100 BC and AD 1600 that were not resolved earlier (Fig. 1a). These excursions are superimposed on a slightly declining long-term trend in $\delta^{13}\text{C}$, which is accompanied by a long-term increase in CH_4 mixing ratios of about 70 p.p.b. between 100 BC and AD 1600 as observed in both Northern Hemisphere and Southern Hemisphere records^{4–7} (Fig. 1b).

We use a two-box model (see Supplementary Information) to infer possible source/sink variations that are consistent with the observed $\delta^{13}\text{C}$ and CH_4 data. In a first step, we vary single sources individually to identify the most important contributors to the isotope variations. The results show that the long-term increase in CH_4 mixing ratios of 70 p.p.b. (corresponding to a source change of 28 Tg CH_4 between 100 BC and AD 1600) cannot originate from geological and pyrogenic sources only, but must be primarily driven by changes in biogenic emissions (Supplementary Fig. 2), in agreement with recent model studies^{1,3}. The three centennial-scale $\delta^{13}\text{C}$ excursions (Fig. 1a) could be caused by either a relative increase in the ^{13}C -enriched sources (pyrogenic or geological sources) or a decrease in the ^{13}C -depleted

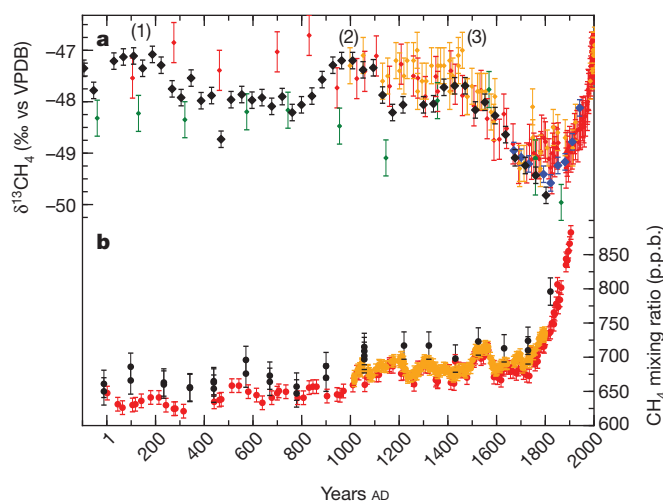


Figure 1 | Records of $\delta^{13}\text{C}$ and mixing ratio of CH_4 over the past two millennia. **a**, $\delta^{13}\text{C}$ measurements on air trapped in Greenland ice cores from NEEM (black diamonds; this study), EUROCORE (blue diamonds; this study), GISPII²³ (green diamonds) and Antarctic ice cores from Law Dome¹ (red diamonds) and the WAIS divide² (orange diamonds). (1), (2) and (3) correspond to the three excursions in the Northern Hemisphere $\delta^{13}\text{C}$ record (see main text). **b**, CH_4 mixing ratio records from Greenland (GRIP^{5,6}; black circles) and Antarctica (Law Dome⁴, red circles; WAIS⁷, orange circles). Each data point represents one measurement. Error bars represent $\pm 1\sigma$, based on the reproducibility of the measurement systems.

¹Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Princetonplein 5, 3584CC Utrecht, The Netherlands. ²ARVE Group, Ecole Polytechnique Fédérale de Lausanne (EPFL), Station 2, 1015 Lausanne, Switzerland. ³Center for Ice and Climate, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen, Denmark. ⁴SRON Netherlands Institute for Space Research, Sorbonnelaan 2, 3584CA Utrecht, The Netherlands. ⁵Earth and Environmental Systems Institute, Penn State University, University Park, Pennsylvania 16802, USA. ⁶UIJF – Grenoble 1/CNRS, Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE), UMR 5183, F-38041 Grenoble, France. ⁷UIJF – Grenoble 1/CNRS, Grenoble Image Parole Signal Automatique (GIPSA-lab), UMR 5216, BP 46, F-38402 St Martin d'Hères, France.

sources (biogenic sources). We consider short-term fluctuations in geologic emissions to be unlikely, so our discussion simplifies to biogenic versus pyrogenic sources. The additional constraint that no clear corresponding signal is observed in the CH_4 mixing ratio implies that the isotope variations cannot be explained without variations in pyrogenic sources (Supplementary Fig. 3). Indeed, the larger the difference between the isotopic signature of specific sources and the global mean, the higher its 'isotope leverage', that is, the more effective a change in this source is in changing the overall isotopic composition.

Whereas this single-source approach is suitable for qualitatively identifying the main drivers of the observed variability, it does not help in quantifying simultaneous changes in multiple sources. Thus, in a second step, we mathematically solve the CH_4 and isotope mass balance equations to determine simultaneous variations in the biogenic and the pyrogenic sources that can explain the measured $\delta^{13}\text{C}$ and CH_4 mixing ratio data (see Supplementary Information). This multiple-source change approach shows that the centennial-scale $\delta^{13}\text{C}$ excursions must be related to increased pyrogenic emissions that are balanced by a concomitant decrease in biogenic emissions.

In order to estimate the uncertainties in the reconstructed source variations, an error propagation study was carried out using a Monte Carlo approach. One thousand simulations were performed, in which model input parameters (observation errors, inter-hemispheric difference in CH_4 and in $\delta^{13}\text{C}$, the geological source strength and isotope signatures of the sources) were perturbed randomly within their range of uncertainty (see Supplementary Information). The resulting source reconstructions fall within the light yellow and light green bands in Fig. 2.

In order to further investigate the origin of the reconstructed variations in CH_4 sources, we compared our pyrogenic and biogenic emission scenarios to the Northern Hemisphere charcoal index¹¹ (based on charcoal accumulation measurements in sediments, and thus a good indicator of biomass burning changes in the past), to model-derived fire activity¹², to the rate of deforestation (estimates based on population data, remote sensed images and land census)^{13,14}, to precipitation estimates¹⁵, and to reconstructions of Northern Hemisphere¹⁶ and extratropical Northern Hemisphere (NEXT)¹⁷ temperatures (Fig. 2). The Northern Hemisphere charcoal index¹¹ (Fig. 2c) shows three peaks in biomass burning between 100 BC and AD 1600 that qualitatively coincide in time with the three excursions in $\delta^{13}\text{C}$, but the overall correlation is weak. The rate of deforestation is a good indicator of anthropogenic variations in biomass burning because fire was the primary means for land maintenance and clearance (see Supplementary Information). We calculated pyrogenic CH_4 emissions from fires used to clear land and to maintain cropland (see Supplementary Information and Supplementary Fig. 7). The results show that between 100 BC and AD 1600, human activity may have been responsible for roughly 20–30% of the total pyrogenic CH_4 emissions. The anthropogenic fraction increases over time, whereas natural pyrogenic emissions decreased. It is important to note that our estimates of the area under land use¹³ include intensification of land use due to new agricultural technologies and population pressure in the last centuries. Therefore the overall land use area increased more slowly than population increase in the recent past. Other estimates of anthropogenic fire activity^{12,14} are based on constant per capita land use area. Using these data, CH_4 would scale directly with population, resulting in lower CH_4 emissions from human activity in the past.

The values of $\delta^{13}\text{C}$ are high at the beginning of our record and decrease around AD 200 (Fig. 1), which results in a decrease of pyrogenic sources and an increase in biogenic sources in our source reconstruction (Fig. 2a, b). Both the Northern Hemisphere charcoal index (Fig. 2c) and the NEXT temperature reconstruction¹⁷ (Fig. 2f) show decreasing values during this period, for which data on fire activity and precipitation are not available. The decrease in NEXT temperatures could have led to lower CH_4 emissions from wildfires, but a causal link to increasing wetland emissions is less obvious. The data on the rate of

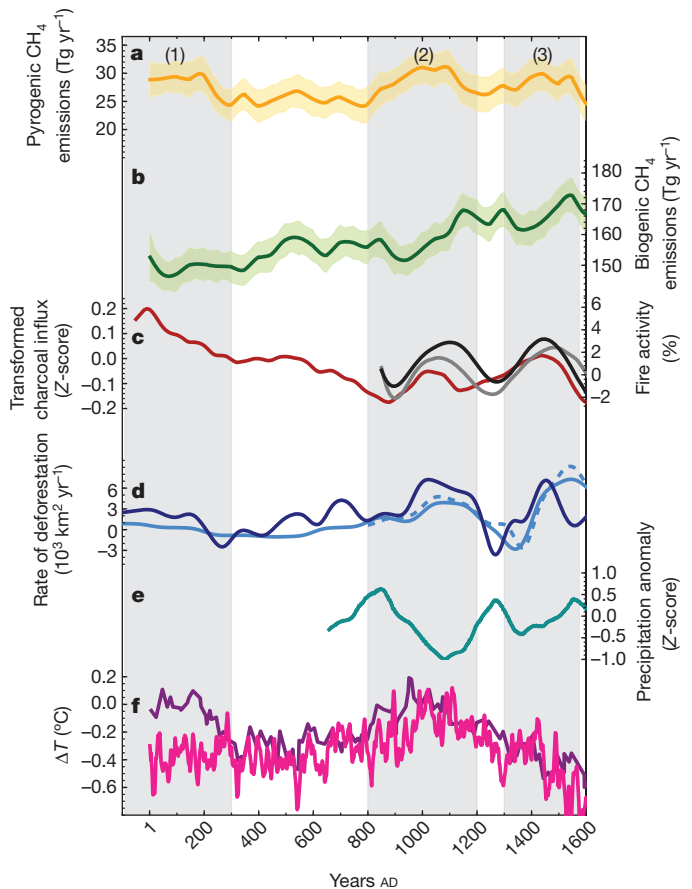


Figure 2 | Reconstructed scenarios for pyrogenic and biogenic CH_4 emissions and other palaeoproxies between 100 BC and AD 1600.

Uncertainty envelopes for the source reconstructions are derived from random parameter perturbations in the source inversion (see Supplementary Information). **a**, **b**, Pyrogenic (**a**) and biogenic (**b**) CH_4 emissions. **c**, Northern Hemisphere transformed charcoal index¹¹ (red curve) and fire activity estimates¹² (model data give percentage of pre-industrial mean) caused by only natural wildfires (grey curve) and by both wildfires and anthropogenic fires (black curve). **d**, Northern Hemisphere rate of deforestation data from KK10¹³ (dark blue), HYDE¹⁴ (light blue solid line) and Pongratz¹⁴ (light blue dashed line). **e**, Northern Hemisphere precipitation anomaly¹⁵. **f**, Northern Hemisphere temperature reconstruction¹⁶ (pink) and Northern Hemisphere extratropics temperature (NEXT) reconstruction¹⁷ (purple). Temperature difference ΔT is with respect to the 1961–90 temperature average. (1), (2) and (3) correspond to the three excursions in the Northern Hemisphere $\delta^{13}\text{C}$ record.

anthropogenic deforestation also show a decrease around AD 200, which is related to drastic population declines in China and Europe following the fall of the Han dynasty and the decline of the Roman empire¹³. Rapidly expanding industrial activity between 100 BC and AD 200 in both Europe and East Asia has been reconstructed from heavy-metal pollution detected in a Greenland ice core¹⁸ and sedimentary records¹⁹. During that time, charcoal was the preferred fuel for industrial and domestic purposes, yielding a large source of ^{13}C -enriched CH_4 (ref. 20). Based on archaeological metal production estimates²¹, we calculate that the charcoal used for metal production at the peak of the Roman empire alone could have produced 0.65 Tg yr^{-1} of CH_4 . Contemporary civilizations in China and India had similar populations and even more sophisticated metal industries²¹. Although specific estimates of metal production are highly uncertain, we suggest that this early industrial activity may have contributed a sizeable fraction to the $\sim 5 \text{ Tg yr}^{-1}$ extra pyrogenic CH_4 emissions before AD 200, according to our model calculation.

The second $\delta^{13}\text{C}$ excursion (Fig. 1) correlates very well in timing and duration with the temperature maximum of the Medieval Climate

Anomaly (MCA) that appears in both the Northern Hemisphere and NEXT temperature reconstructions (Fig. 2f) between AD 800 and AD 1200. During this time, our reconstructed scenarios show a temporary decrease in biogenic sources and an increase in pyrogenic sources (Fig. 2). Widespread extended droughts¹⁵ associated with precipitation decrease in Northern Europe (Fig. 2e) provided favourable conditions for enhanced wildfire activity during the MCA. In addition, the medieval period was a time of accelerating deforestation (Fig. 2d) as a result of population expansion and urbanization in Europe and Asia¹³, which coincides in time with the second $\delta^{13}\text{C}$ excursion observed. Fire activity data (Fig. 2c) provide independent support for both natural and anthropogenic contributions to the increased pyrogenic emissions during this period.

Interestingly, during the MCA the isotopic composition of CH_4 implies generally decreasing biogenic emissions. This seems to contradict the idea that high latitude Northern Hemisphere wetland CH_4 emissions acted as a positive climate feedback to increasing Northern Hemisphere temperatures, which would have led to a decrease in $\delta^{13}\text{C}$, contrary to our observations. Increases in methanogenesis rates caused by higher temperatures during this period may have been compensated by decreases in wetland area due to extended droughts²², leading to a net decrease in biogenic CH_4 emissions.

The third maximum in $\delta^{13}\text{C}$ (Fig. 1) occurs simultaneously with the onset of the Little Ice Age (LIA). During this period, Northern Hemisphere temperatures and precipitation decreased (Fig. 2e, f). This may have led to unfavourable conditions for enhanced biogenic emissions. Increased fire activity and charcoal index values during the same period indicate that CH_4 from natural wildfires may have made a significant contribution to the observed $\delta^{13}\text{C}$ excursion. In addition, the rate of deforestation (Fig. 2d) shows an increase that is slightly delayed relative to the third $\delta^{13}\text{C}$ excursion, suggesting more pyrogenic emissions caused by rapid land clearance during this period.

The long-term trend in mixing ratios and $\delta^{13}\text{C}$ values of CH_4 from biogenic sources between 100 BC and AD 1600 (Fig. 1b) can have natural (primarily wetlands^{1,6,23}) and anthropogenic (primarily agricultural^{2,3,24}) components. Recent model calculations⁸ inferred an orbitally-controlled late Holocene increase in global CH_4 levels, primarily driven by increases of Southern Hemisphere tropical natural wetland emissions, which according to this theory are linked to variability in monsoon patterns during the Holocene²⁵. On the other hand, we show in Fig. 3 that the long-term trend in CH_4 mixing ratio—that is, the increase between

100 BC and AD 1800—is in very good agreement with reconstructed global anthropogenic land use¹³. This suggests that human activities, including the expansion of rice agriculture²⁶, played an important role in the observed long-term CH_4 trend over the past two millennia.

Our new isotope data from air trapped in Greenland ice cores allow the reconstruction of variations in different CH_4 source categories over the past 2,100 years. The changes seen in our $\delta^{13}\text{C}$ ice core record cannot be explained without variability in biomass burning, which correlates qualitatively with the charcoal index and fire activity data. In addition, we show that the reconstructed source variations are correlated with anthropogenic activities, in particular with long-term increases in agricultural emissions and with varying levels of biomass burning during the period of the Roman empire and the Han dynasty, the MCA and the onset of the LIA. It is thus likely that human activity contributed to variations in CH_4 emissions to the atmosphere long before pre-industrial times.

METHODS SUMMARY

We analysed 47 ice samples from the NEEM ice core (North Greenland: site coordinates 77° 27' N, 51° 4' W) and 9 from the EUROCORE (Summit, Central Greenland: site coordinates: 72° 34' N, 38° 27' W) using a recently developed dry extraction procedure²⁷ followed by isotope ratio mass spectrometry. A layer of ice (3–5 mm) was microtomed from the surface of the ice to exclude contamination from the drilling liquid and possible anomalies caused by post-coring processes. Samples were not measured in chronological order. Contamination due to sample handling during coring and processing of the ice is very unlikely, because all samples were handled in the same manner and the observed $\delta^{13}\text{C}$ variations follow systematic patterns rather than showing an erratic behaviour. No similar temporal patterns are found in the recently obtained dust and ion records of the NEEM core (M. Bigler, personal communication), which makes *in situ* CH_4 formation in the ice also unlikely. Moreover, *in situ* production is very likely to be biogenic and would lead to increases of isotopically depleted CH_4 , which are not observed in our profile.

The precision of the $\delta^{13}\text{C}$ measurements, based on standards and replicate ice core measurements, is 0.12‰. All $\delta^{13}\text{C}$ values are reported versus VPDB (Vienna Pee Dee Belemnite). We corrected our data for an inter-laboratory offset of 0.51‰ between our laboratory and Pennsylvania State University^{2,23}, as determined in a recent intercalibration exercise using real ice core samples²⁷. In addition, all isotope data were corrected for gravitational settling, and the four shallowest EUROCORE data were corrected for diffusive fractionation²⁸. The age of each ice core air sample was calculated using a delta age (gas age minus ice age) of 183 years for NEEM²⁹ and 210 years for EUROCORE³⁰. Modelled gas age distributions of NEEM and EUROCORE are presented in Supplementary Information.

Received 24 December 2011; accepted 31 July 2012.

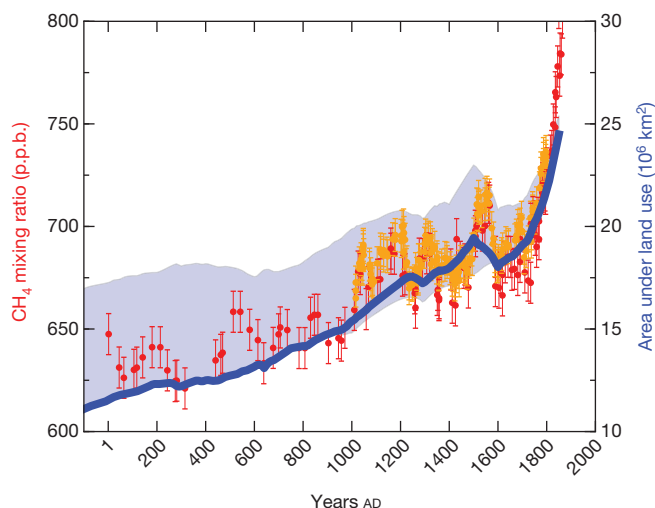


Figure 3 | Comparison between estimate of area under land use and CH_4 mixing ratio. CH_4 mixing ratio records are from the Law Dome⁴ (red circles) and WAIS⁷ (orange circles) ice cores. The shaded area represents an uncertainty estimate on global area under land use (blue curve, from ref. 13). For details on methods used to construct the land use scenario, see Supplementary Information section 3.

1. Ferretti, D. *et al.* Unexpected changes to the global methane budget over the past 2000 years. *Science* **309**, 1714–1717 (2005).
2. Mischler, J. A. *et al.* Carbon and hydrogen isotopic composition of methane over the last 1000 years. *Glob. Biogeochem. Cycles* **23**, GB4024, <http://dx.doi.org/10.1029/2009GB003460> (2009).
3. Houweling, S. *et al.* Early anthropogenic CH_4 emissions and the variation of CH_4 and $^{13}\text{CH}_4$ over the last millennium. *Glob. Biogeochem. Cycles* **22**(1), <http://dx.doi.org/10.1029/2007GB002961> (2008).
4. MacFarling Meure, C. *et al.* Law Dome CO_2 , CH_4 and N_2O ice core records extended to 2000 years BP. *Geophys. Res. Lett.* **33**, L14810 <http://dx.doi.org/10.1029/2006GL026152> (2006).
5. Blunier, T. *et al.* Variations in atmospheric methane concentration during the Holocene epoch. *Nature* **374**, 46–49 (1995).
6. Chappellaz, J. *et al.* Changes in atmospheric CH_4 gradient between Greenland and Antarctica during Holocene. *J. Geophys. Res.* **102** (D13), 15987–15997 (1997).
7. Mitchell, L. E. *et al.* Multidecadal variability of atmospheric methane, 1000–1800 CE. *J. Geophys. Res.* **116**, G02007, <http://dx.doi.org/10.1029/2010JG001441> (2011).
8. Singarayer, J. S. *et al.* Late Holocene methane rise caused by orbitally controlled increase in tropical sources. *Nature* **470**, 82–85 (2011).
9. Quay, P. *et al.* The isotopic composition of atmospheric methane. *Glob. Biogeochem. Cycles* **13**, 445–461 (1999).
10. Etiope, G., Lassey, K. R., Klusman, R. W. & Boschi, E. Reappraisal of the fossil methane budget and related emission from geologic sources. *Geophys. Res. Lett.* **35**, L09307, <http://dx.doi.org/10.1029/2008GL033623> (2008).
11. Marlon, J. R. *et al.* Climate and human influences on global biomass burning over the past two millennia. *Nature Geosci.* **1**, 697–702 (2008).
12. Pechony, O. & Shindell, D. T. Driving forces of global wildfires over the past millennium and the forthcoming century. *Proc. Natl Acad. Sci. USA* **107**, 19167–19170 (2010).

13. Kaplan, J. O. *et al.* Holocene carbon emissions as a result of anthropogenic land cover change. *Holocene* **21**, 775–791 (2011).
14. Schmidt, G. A. *et al.* Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.1). *Geosci. Model Dev.* **5**, 185–191 (2012).
15. Helama, S., Merilainen, J. & Tuomenvirta, H. Multicentennial megadrought in northern Europe coincided with a global El Niño-Southern Oscillation drought pattern during the Medieval Climate Anomaly. *Geology* **37**, 175 (2009).
16. Moberg, A. *et al.* Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature* **433**, 613–617 (2005).
17. Ljungqvist, F. C. A new reconstruction of temperature variability in the extra-tropical northern hemisphere during the last two millennia. *Geogr. Annal. A* **92**, 339–351 (2010).
18. Hong, S. M., Candelone, J. P., Patterson, C. C. & Boutron, C. F. History of ancient copper smelting pollution during Roman and medieval times recorded in Greenland ice. *Science* **272**, 246–249 (1996).
19. Brännvall, M. L. *et al.* The Medieval metal industry was the cradle of modern large scale atmospheric lead pollution in northern Europe. *Environ. Sci. Technol.* **33**, 4391–4395 (1999).
20. Akagi, S. K. *et al.* Emission factors for open and domestic biomass burning for use in atmospheric models. *Atmos. Chem. Phys.* **11**, 4039–4072 (2011).
21. Craddock, P. T. *Handbook of Engineering and Technology in the Classical World* Ch. 4 98–120 (Oxford Univ. Press, 2008).
22. Ringeval, B. *et al.* Climate-CH₄ feedback from wetlands and its interaction with the climate-CO₂ feedback. *Biogeosciences* **8**, 2137–2157 (2011).
23. Sowers, T. Atmospheric methane isotope records covering the Holocene period. *Quat. Sci. Rev.* **29**, 213–221 (2010).
24. Ruddiman, W. F. & Thomson, J. S. The case for human causes of increased atmospheric CH₄. *Quat. Sci. Rev.* **20**, 1769–1777 (2001).
25. Burns, S. J. Speleothem records of changes in tropical hydrology over the Holocene and possible implications for atmospheric methane. *Holocene* **21**, 735–741 (2011).
26. Fuller, D. Q. *et al.* The contribution of rice agriculture and livestock pastoralism to prehistoric methane levels: an archaeological assessment. *Holocene* **21**, 743–759 (2011).
27. Sapart, C. J. *et al.* Simultaneous stable isotope analysis of methane and nitrous oxide on ice core samples. *Atmos. Meas. Tech.* **4**, 2607–2618 (2011).
28. Sapart, C. J. *et al.* Reconstruction of the carbon isotopic composition of methane over the last 50 yr based on firn air measurements at 11 polar sites. *Atmos. Chem. Phys. Discuss.* **12**, 9587–9619 (2012).
29. Buizert, C. *et al.* Gas transport in firn: multiple-tracer characterisation and model intercomparison for NEEM, Northern Greenland. *Atmos. Chem. Phys. Discuss.* **11**, 15975–16021 (2011).
30. Schwander, J. *et al.* The age of the air in the firn and the ice at Summit, Greenland. *J. Geophys. Res.* **98** (D2), 2831–2838 (1993).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Bigler for sharing NEEM CFA dust and ion data; the NEEM community for providing us with ice core samples; and O. Pechony, J. Marlon and Z. Wang for sharing data on fire activity and the charcoal index. This project was supported by the Dutch Science Foundation (NOW; projects 851.30.020 and 865.07.001). NEEM is directed and organized by the Center of Ice and Climate at the Niels Bohr Institute and the US NSF, Office of Polar Programs; it is supported by funding agencies and institutions in Belgium (FNRS-CFB and FWO), Canada (NRCan/GSC), China (CAS), Denmark (FIST), France (IPEV, CNRS/INSU, CEA and ANR), Germany (AWI), Iceland (Rannls), Japan (NIPR), Korea (KOPRI), The Netherlands (NWO/ALW), Sweden (VR), Switzerland (SNF), the UK (NERC) and the USA (US NSF, Office of Polar Programs, ARC0806407). J.O.K. and K.M.K. were supported by the Swiss National Science Foundation (grant PP0022_119049) and FIRB project CASTANEA (RBID08LNFJ).

Author Contributions C.J.S., T.R., R.S.W.v.d.W., J.O.K. and G.M. wrote the manuscript. T.R. and R.S.W.v.d.W. planned and designed the study. C.J.S., T.R., R.S.W.v.d.W., J.O.K., G.M., S.H., M.C.K., K.M.K., P.S., T.S., M.P. and T.B. worked on the scientific interpretation. C.J.S., M.P. and C.v.d.V. carried out the measurements. G.M., C.J.S., S.H., P.M. and E.W. carried out the modelling work. J.O.K. and K.M.K. provided reconstructions of human land use. D.D.-J. and T.B. led and coordinated the NEEM ice core drilling project and gas consortium.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.J.S. (c.j.sapart@uu.nl).

A Silurian armoured aplacophoran and implications for molluscan phylogeny

Mark D. Sutton¹, Derek E. G. Briggs², David J. Siveter³, Derek J. Siveter^{4,5} & Julia D. Sigwart⁶

The Mollusca is one of the most diverse, important and well-studied invertebrate phyla; however, relationships among major molluscan taxa have long been a subject of controversy^{1–9}. In particular, the position of the shell-less vermiform Aplacophora and its relationship to the better-known Polyplacophora (chitons) have been problematic: Aplacophora has been treated as a paraphyletic or monophyletic group at the base of the Mollusca^{3,6,8}, proximate to other derived clades such as Cephalopoda^{2,3,10}, or as sister group to the Polyplacophora, forming the clade Aculifera^{1,5,7,11,12}. Resolution of this debate is required to allow the evolutionary origins of Mollusca to be reconstructed with confidence. Recent fossil finds^{13–16} support the Aculifera hypothesis, demonstrating that the Palaeozoic-era palaeoloricate ‘chitons’ included taxa combining certain polyplacophoran and aplacophoran characteristics⁵. However, fossils combining an unambiguously aplacophoran-like body with chiton-like valves have remained elusive. Here we describe such a fossil, *Kulindroplax perissokomos* gen. et sp. nov., from the Herefordshire Lagerstätte^{17,18} (about 425 million years BP), a Silurian deposit preserving a marine biota¹⁸ in unusual three-dimensional detail. The specimen is reconstructed three-dimensionally through physical–optical tomography¹⁹. Phylogenetic analysis indicates that this and many other palaeoloricate chitons are crown-group aplacophorans.

Phylum Mollusca

Kulindroplax perissokomos Sutton, Briggs, Siveter, Siveter and Sigwart gen. et sp. nov.

Etymology. *kulindros* (Greek; ‘roller’ or ‘cylinder’) plus *plax* (Greek; ‘plate’), alluding to the cylindrical valve-bearing morphology; and *perissokomos* (Greek; ‘exceedingly hairy’), alluding to the pervasive spicule covering. Gender masculine.

Holotype. Oxford University Museum of Natural History OUMNH C.29641, the only known specimen (Fig. 1a–m).

Locality and horizon. Wenlock Series, Silurian; Herefordshire, England.

Diagnosis for genus and species. Series of seven similar unarticulated valves, lacking ventral articulamentum. The head valve is shorter and much lower in height; the tail valve is taller. Valves with weak anterior jugal embayment, weakly obtuse apical angle and jugal angle near 90°. Posterolateral margins are weakly concave and serrated. All valves are mixoperipheral, with ventral apical area extending to over 25% of valve length. Valves lack ornament. The foot is absent and the girdle complete ventrally while open anteriorly and posteriorly; it bears concavo-convex blade-like spicules. A gill array of at least four pairs of elements is contained within the long posterior cavity.

Description. The trunk is elongate, moderately curved ventrally (Fig. 1e) and bears seven morphologically similar valves (Fig. 1a, b, e) that overlap without articulating; these would have been in contact

when the trunk was straight. In each valve, width > length > height (see Supplementary Note 1 for biometrics). Valves II–VI differ only slightly, but valve I is relatively short and low and valve VII (posterior terminal valve) is substantially taller (Fig. 1a, b, g, h). In transverse section the valves are keeled posteriorly and rounded anteriorly, with planar to weakly convex lateral areas and a jugal angle of approximately 90° (Fig. 1g), except in valve I (approximately 120°). In lateral profile (Fig. 1e), the dorsal surface is weakly convex, the anterolateral margin is evenly curved and the posterolateral margin is straight to weakly concave. In dorsoventral views (Fig. 1a, b), the anterior margin shows a weak median jugal embayment. Lateral margins are straight to weakly convex, and anterolateral corners are well rounded. Posterolateral corners are sharp in posterior valves and more rounded in anterior valves, with angles increasing anteriorly from approximately 80° (valve VII) to approximately 120° (valve I). Posterolateral margins are weakly concave, converging to a rounded beak with a weakly obtuse apical angle. Weakly defined linear depressions on the lateral areas of valves converge posteriorly at the apex (for example, Fig. 1a, valve V). Posterolateral margins are weakly serrated (Fig. 1i). No growth lines or dorsal ornament have been observed. There is a mixoperipheral fold around the posterior margin of all valves on the ventral internal surface (apical area, Fig. 1b, k). Valves are up to 0.4 mm thick away from apical areas.

The girdle forms a cuticular cylinder 10.5 mm wide and 38 mm long (excluding valves and spicules), with open subplanar terminations (Fig. 1e, h); the anterior termination is flared ventrally. A foot is absent; a weak ventral ridge consisting of imperisistently preserved thickening of ventromedial cuticle extends to both terminations (Fig. 1c, h, l). The cuticle is attached to the ventral part of the valves near their lateral margins (Fig. 1l) from anterior of apical area to near the posterolateral corners, except in the anteriormost valve in which attachment (preserved on right only, Fig. 1h) does not extend anterior of mid-length. Where attached, the cuticle displays reflexion (laterally concave); between attachments it is subcircular in transverse section (Fig. 1c, l; note that Fig. 1c is an oblique section). Densely packed spicules project subnormally from the entire girdle surface except where it attaches to valves, and along the ventral cuticular ridge (Fig. 1c–e, h, j, m). Spicules are blade-like with a sub-semicircular cross-section (Fig. 1d); they are convex anteriorly and weakly recurved to point posteriorly (Fig. 1j, m). Spicules are 1–2 mm long, with larger spicules occurring laterally. Spicules taper to pointed tips (Fig. 1m); some evenly, others maintaining maximum breadth for much of their length.

The body mass is shrunken owing to decay and appears undifferentiated; valves III–VII preserve the attachment (Fig. 1b). Body mass approaches (but is not attached to) the ventral cuticle only near valve V; elsewhere it slopes dorsally. The surface of mass is uneven, asymmetrical and marked by subcircular ‘pockmarks’ (for example, Fig. 1k). An anterior dorsomedian linear depression may represent the position of the gut (Fig. 1b). No radula is preserved.

¹Department of Earth Sciences and Engineering, Imperial College London, London SW7 2BP, UK. ²Department of Geology & Geophysics, and Yale Peabody Museum of Natural History, Yale University, PO Box 208109, New Haven, Connecticut 06520-8109, USA. ³Department of Geology, University of Leicester, Leicester LE1 7RH, UK. ⁴Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3AN, UK. ⁵Geological Collections, University Museum of Natural History, Oxford OX1 3PW, UK. ⁶Queen’s University Belfast, School of Biological Sciences, Marine Laboratory, Portaferry, Northern Ireland BT22 1PF, UK.

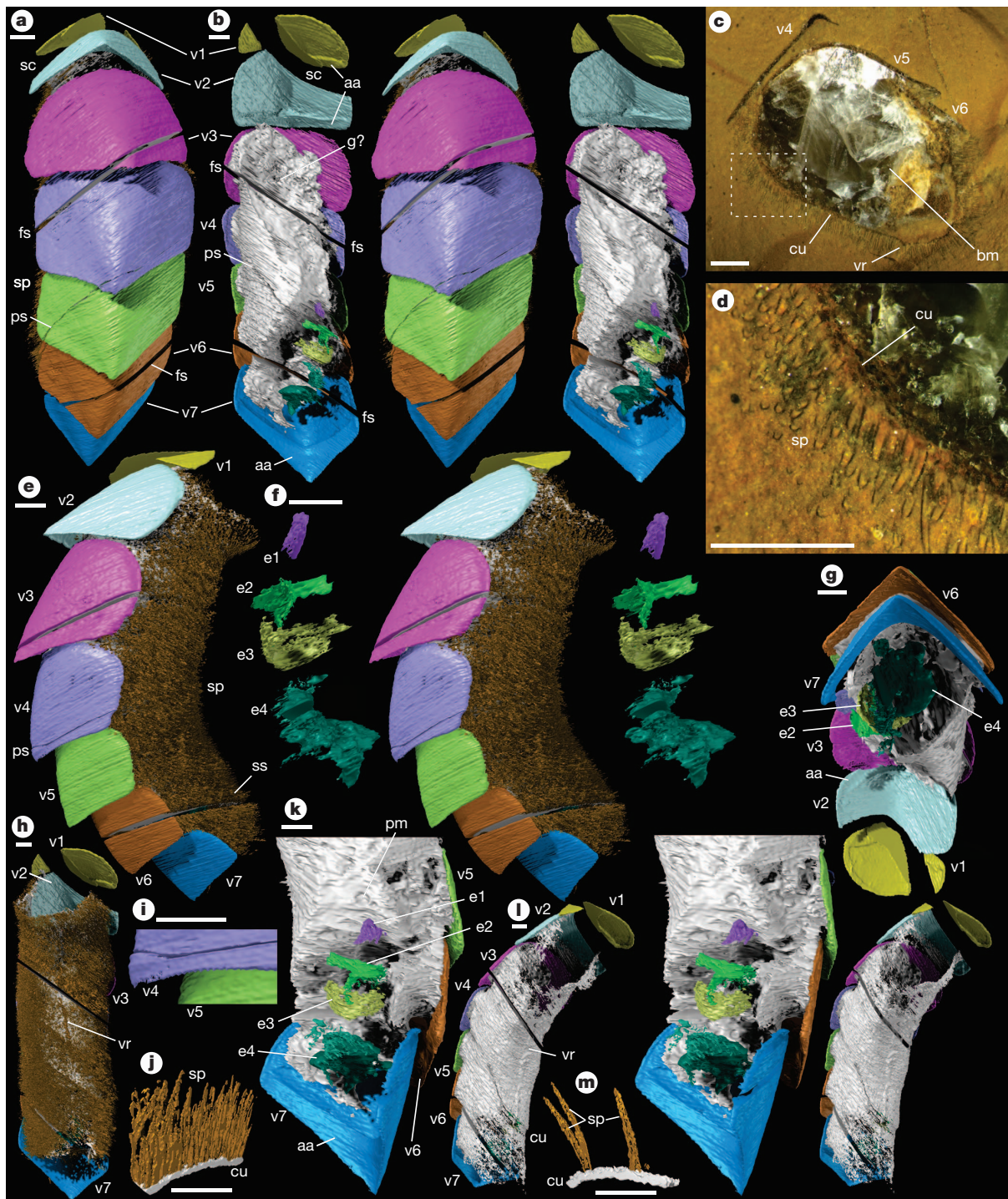


Figure 1 | Oxford University Museum of Natural History (OUMNH) C.29641: holotype of *Kulindroplax perissokomos*. **a**, **b** and **e–m** are 'virtual' reconstructions. **a**, Dorsal stereo pair. **b**, Ventral stereo pair (cuticle and spicules removed). **c**, Photograph of the specimen before serial grinding (that is, section along primary split). Dashed box indicates the position of **d**. **d**, Detail of **c** showing spicules. **e**, Lateral stereo pair. **f**, Dorsal stereo pair of gill array (cuticle, spicules, valves and body removed). **g**, Posterior view (cuticle and spicules removed). **h**, Ventral view. **i**, Detail of **e** (rotated) showing serrated margin of valve IV. **j**, Representative piece of spicule-bearing cuticle, from point labelled 'ss' in **e**. Posterolateral view (ventral up). **k**, Ventrolateral view of

posterior of specimen (cuticle and spicules removed). **l**, Oblique (subventral) view (spicules removed). **m**, Sub-posterior view (ventral upright) of cuticle piece from **j** with all but three spicules removed. Scale bars, 2 mm (virtual reconstructions are perspective views and scale decreases away from viewer; where depth of object is substantial, the scale is calculated for the valve closest to the viewer). aa, apical area; bm, body mass; cu, cuticle; e1–e4, elements 1–4 of gill array; fs, fine-cut (~0.3 mm material removed); g, gut trace; pm, 'pockmark'; ps, primary split (crack between part and counterpart); sc, saw-cut (~2 mm material removed); sp., spicules; ss, spicule sample; v1–v7, valves I (head)–VII (tail); vr, ventral ridge.

A posterior cavity contains poorly preserved filament-bearing structures that are interpreted as elements of a gill array comprising four element pairs (Fig. 1b, f, g, k). These are described in detail in Supplementary Note 3.

Discussion. The morphology of *Kulindroplax* provides new and clear documentation of the flexibility and disparity of the palaeoloricate bauplan^{5,13–16}, as well as supporting a natural clade Aculifera. Two other Palaeozoic fossils show both polyplacophoran- and aplacophoran-like

characteristics: *Acaenoplax*^{13,14}, also from the Herefordshire Lagerstätte, and the less well-preserved Ordovician *Phthipodochiton thraivensis*^{5,16}. The arrangement and morphology of the valves of *Acaenoplax* differ from those in living polyplacophorans and their fossil representatives, and some features of the body have no direct equivalent in living aplacophorans^{13,14,20,21}. The valves of *Phthipodochiton* in contrast, are typical of palaeoloricate ‘polyplacophorans’; this taxon also has a ventrally complete (or near-complete) girdle, although this feature is not sufficiently well preserved for satisfactory characterization. *Kulindroplax* represents the first unambiguous combination of palaeoloricate valves and aplacophoran body, and hence is the clearest known fossil link between the Polyplacophora and Aplacophora. The valves of *Kulindroplax* are closely comparable to those of the palaeoloricate *Chelodes*²², and of *Phthipodochiton*. However, the cuticle, sub-circular cross-section, complete coat of spicules and wide posterior cavity housing respiratory organs are aplacophoran features. *Phthipodochiton* has been compared with predatory solenogasters (neomeniomorphs)^{5,23}. Several characters of *Kulindroplax*, in contrast, invite comparison with the caudofoveate (chaetoderm) aplacophorans. These include the absence of a pedal groove or pit, the posteriorly (as opposed to posteroventrally) opening respiratory cavity, and the gill array with elements that are more closely comparable to the ctenidia of caudofoveates than the gill folds of the solenogasters, although there seem to be four pairs rather than one.

All valve-bearing aculiferans possess eight dorsal valves where this number can be determined, except for *Acaenoplax*^{13,14} and Multiplacophora¹⁵. *Kulindroplax* is unique in its chiton-like scleritome (that is, a single row of mixoperipherally grown overlapping dorsal valves) with only seven elements (see Supplementary Note 4, character 14), although there is a parallel in the seven putative shell-fields described in one postlarval solenogaster²⁴.

The phylogenetic position of *Kulindroplax* was investigated using a modified version of the data matrix described in ref. 5. The a priori weighting scheme of these authors is replaced here with a posteriori weighting schemes^{25,26}, although results under a priori weighting are similar. Supplementary Note 4 provides full details of the new matrix, the methods used and the results of different variants of our analysis. Our tree topology (Fig. 2) is similar to that depicted in ref. 5, recovering clades including the halwaxiids²⁷ (*Wiwaxia*, *Halkieria* and *Orthrozacnus*), Neoloricata and Aculifera (Aplacophora and Polyplacophora). Aculifera falls out as sister group to Conchifera, the latter represented by the monoplacophoran *Neopilina*, the early crown-group bivalves *Pojetaia* and *Fordilla*, and an Ordovician bivalve (*Babinka*) that had eightfold iteration of adductor muscles²⁸. The recovery of Aculifera and Conchifera in all variants of our analysis demonstrates the strong morphological support that fossil data now offer for the Aculifera

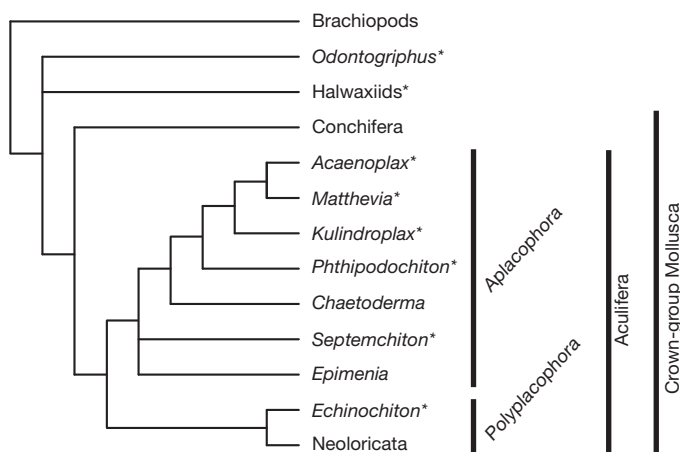


Figure 2 | Summary of parsimony analysis results. Asterisk denotes extinct taxa. See Supplementary Note 4 for details and full version of tree.

hypothesis, paralleling recent molecular evidence favouring this topology^{1,2,11,12}. The Serialia hypothesis for a clade comprising monoplacophorans and polyplacophorans^{2,3} is rejected by present fossil data.

Aplacophora (represented by *Chaetoderma* and *Epimenia*) is also recovered, but instead of occupying a derived position relative to related valve-bearing fossils⁵, the extant aplacophoran genera occur relatively stemwards. *Epimenia* (a solenogaster) resolves as more primitive than *Chaetoderma* (a caudofoveate). This result is compatible with neontological inferences, developed under the Aculifera hypothesis, that caudofoveates are relatively derived⁷.

Evidence for the mode of life of *Acaenoplax*¹⁴ and *Phthipodochiton*²³ suggests a creeping epifaunal habit, but the concavo-convex spicules of *Kulindroplax* seem to be well adapted as sediment ratchets that could have facilitated forward movement through a substrate. Although these spicules occur all over the cuticle, the dorsal valves are not easy to reconcile with a fully infaunal (caudofoveate-like) autecology; *Kulindroplax* may instead have been semi-infaunal, probably facultatively, retaining a requirement for dorsal armour.

Kulindroplax, *Acaenoplax*, *Phthipodochiton* and other palaeoloricates fall within the aplacophoran crown-group in the results from all but one of our analyses (the exception neither recovers nor excludes this position; see Supplementary Note 4). Thus, the presence of a foot cannot be inferred for any animal with chiton-like palaeoloricate valves. These topologies also indicate that valve loss occurred independently in Solenogastres and Caudofoveata: it is not a synapomorphy of Aplacophora. Aplacophorans were primitively shelled molluscs; the few living forms represent the survivors of a diverse valve-bearing Palaeozoic clade.

METHODS SUMMARY

Morphology was reconstructed digitally following serial grinding at 30-μm intervals; compressed versions of the raw data are available for download as Supplementary Data 2. Some material has been lost to saw-cuts in processing; see Supplementary Note 2. The full reconstructed model is available for download and viewing in VAXML format²⁹ as Supplementary Data 1. See Supplementary Note 4 for details of the phylogenetic analyses.

Received 29 February 2012; accepted 14 June 2012.

- Kocot, K. M. *et al.* Phylogenomics reveals deep molluscan relationships. *Nature* **477**, 452–456 (2011).
- Wilson, N. G., Rouse, G. W. & Giribet, G. Assessing the molluscan hypothesis Serialia (Monoplacophora + Polyplacophora) using novel molecular data. *Mol. Phylogenet. Evol.* **54**, 187–193 (2010).
- Giribet, G. *et al.* Evidence for a clade composed of molluscs with serially repeated structures: monoplacophorans are related to chitons. *Proc. Natl Acad. Sci. USA* **103**, 7723–7728 (2006).
- Passamaneck, Y. J., Schander, C. & Halanych, K. M. Investigation of molluscan phylogeny using large-subunit and small-subunit nuclear rRNA sequences. *Mol. Phylogenet. Evol.* **32**, 25–38 (2004).
- Sigwart, J. D. & Sutton, M. D. Deep molluscan phylogeny: synthesis of palaeontological and neontological data. *Proc. R. Soc. B* **274**, 2413–2419 (2007).
- Salvini-Plawen, L. & Steiner, G. in *Origin and Evolutionary Radiation of the Mollusca* (ed. Taylor, J. D.), Ch. 2, 29–51 (Oxford Univ. Press, 1996).
- Schelteima, A. H. Aplacophora as progenetic aculiferans and the coelomate origin of mollusks as the sister taxon of Sipuncula. *Biol. Bull.* **184**, 57–78 (1993).
- Salvini-Plawen, L. On the phylogenetic significance of the aplacophoran Mollusca. *Iberus* **21**, 67–97 (2003).
- Telford, M. J. & Budd, G. E. Invertebrate evolution: bringing order to the molluscan chaos. *Curr. Biol.* **21**, R964–R966 (2011).
- Meyer, A., Wittek, A. & Lieb, B. Selecting ribosomal protein genes for invertebrate phylogenetic inferences: how many genes to resolve the Mollusca? *Methods Ecol. Evol.* **2**, 34–42 (2011).
- Vinther, J., Sperling, E. A., Briggs, D. E. G. & Peterson, K. A molecular palaeobiological hypothesis for the origin of aplacophoran molluscs and their derivation from chiton-like ancestors. *Proc. R. Soc. B* **279**, 1259–1268 (2012).
- Smith, S. A. *et al.* Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* **480**, 364–367 (2011).
- Sutton, M. D., Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. An exceptionally preserved vermiform mollusc from the Silurian of England. *Nature* **410**, 461–463 (2001).
- Sutton, M. D., Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. Computer reconstruction and analysis of the vermiform mollusc *Acaenoplax hayae* from the Herefordshire Lagerstätte (Silurian, England), and implications for molluscan phylogeny. *Palaeontology* **47**, 293–318 (2004).

15. Vendrasco, M. J., Wood, T. E. & Runnegar, B. N. Articulated Palaeozoic fossil with 17 plates greatly expands disparity of early chitons. *Nature* **429**, 288–291 (2004).
16. Sutton, M. D. & Sigwart, J. D. A chiton without a foot. *Palaeontology* **55**, 401–411 (2012).
17. Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. Soft-bodied fossils from a Silurian volcanoclastic deposit. *Nature* **382**, 248–250 (1996).
18. Briggs, D. E. G., Siveter, David J., Siveter, Derek J. & Sutton, M. D. Virtual fossils from a 425 million-year-old volcanic ash. *Am. Sci.* **96**, 474–481 (2008).
19. Sutton, M. D. Tomographic techniques for the study of exceptionally preserved fossils. *Proc. R. Soc. B* **275**, 1587–1593 (2008).
20. Steiner, G. & Salvini-Plawen, L. *Acaenoplax* — polychaete or mollusc? *Nature* **414**, 601 (2001).
21. Sutton, M. D., Briggs, D. E. G., Siveter, David J. & Siveter, Derek J. *Acaenoplax* — polychaete or mollusc? (response to Steiner and Salvini-Plawen). *Nature* **414**, 602 (2001).
22. Cherns, L. *Chelodes* and closely related Polyplacophora (Mollusca) from the Silurian of Gotland, Sweden. *Palaeontology* **41**, 545–573 (1998).
23. Donovan, S. K., Sutton, M. D. & Sigwart, J. D. Crinoids for lunch? An unexpected biotic interaction from the Upper Ordovician of Scotland. *Geology* **38**, 935–938 (2010).
24. Scheltema, A. H. & Ivanov, D. L. An aplacophoran postlarva with iterated dorsal groups of spicules and skeletal similarities to Paleozoic fossils. *Invertebr. Biol.* **121**, 1–10 (2002).
25. Farris, J. S. A successive approximations approach to character weighting. *Syst. Zool.* **18**, 374–385 (1969).
26. Goloboff, P. A. Estimating character weights during tree search. *Cladistics* **9**, 83–91 (1993).
27. Conway Morris, S. & Caron, J.-B. Halwaxiids and the early evolution of the lophotrochozoans. *Science* **315**, 1255–1258 (2007).
28. McAlester, A. L. Systematics, affinities, and the life habits of *Babinka*, a traditional Ordovician lucinoid bivalve. *Palaeontology* **8**, 231–246 (1965).
29. Sutton, M. D., Garwood, R. J., Siveter, David J. & Siveter, Derek J. SPIERS and VAXML; a software toolkit for tomographic visualisation and a format for virtual specimen interchange. *Palaeontologia Electronica* **15**, <http://palaeo-electronica.org/content/94-issue-2-2012-technical-articles/226-virtual-palaeontology-toolkit> (2012).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The Natural Environmental Research Council (NERC Grant NE/F018037/1) supported this research. K. Saunders, K. Davies, C. Dean, D. Legg and C. Lewis provided technical assistance, and the staff of Tarmac Western and the late R. Fenn facilitated the fieldwork.

Author Contributions D.J.S., D.J.S., D.E.G.B. and M.D.S. carried out fieldwork. M.D.S. and J.D.S. performed phylogenetic analyses. M.D.S. reconstructed the specimen and wrote the paper, with scientific and editorial input from all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.D.S. (m.sutton@imperial.ac.uk).

A transcriptomic hourglass in plant embryogenesis

Marcel Quint¹, Hajk-Georg Drost², Alexander Gabel², Kristian Karsten Ullrich¹, Markus Bönn^{2,3} & Ivo Grosse²

Animal and plant development starts with a constituting phase called embryogenesis, which evolved independently in both lineages¹. Comparative anatomy of vertebrate development—based on the Meckel-Serrès law² and von Baer's laws of embryology³ from the early nineteenth century—shows that embryos from various taxa appear different in early stages, converge to a similar form during mid-embryogenesis, and again diverge in later stages. This morphogenetic series is known as the embryonic 'hourglass'^{4,5}, and its bottleneck of high conservation in mid-embryogenesis is referred to as the phylotypic stage⁶. Recent analyses in zebrafish and *Drosophila* embryos provided convincing molecular support for the hourglass model, because during the phylotypic stage the transcriptome was dominated by ancient genes⁷ and global gene expression profiles were reported to be most conserved⁸. Although extensively explored in animals, an embryonic hourglass has not been reported in plants, which represent the second major kingdom in the tree of life that evolved embryogenesis. Here we provide phylotranscriptomic evidence for a molecular embryonic hourglass in *Arabidopsis thaliana*, using two complementary approaches. This is particularly significant because the possible absence of an hourglass based on morphological features in plants suggests that morphological and molecular patterns might be uncoupled. Together with the reported developmental hourglass patterns in animals, these findings indicate convergent evolution of the molecular hourglass and a conserved logic of embryogenesis across kingdoms.

In flowering plants, embryogenesis can be separated into three major phases. The early phase is characterized by asymmetric cell divisions to establish apical–basal polarity. In the intermediate phase, major organs and primordia are initiated, which expand in the late phase to the mature embryo^{9,10}. One notable difference between embryogenesis in animals and plants concerns the establishment of morphological variation between taxa. For example, vertebrates develop morphological variation in late embryogenesis, whereas differences between flowering plant taxa are only established during post-embryonic development. Inspired by the historical relevance of the embryonic hourglass model in animals, by recent transcriptional support from studies in zebrafish⁷ and *Drosophila*^{7,8}, and by the absence of any reported anatomical evidence for such a pattern during plant embryogenesis, we assess the possible existence of a transcriptional hourglass during embryogenesis of the plant reference species *A. thaliana*.

Recently, genome-wide expression profiles of a complete developmental series from the zygote to the mature embryo in *A. thaliana* were obtained¹¹. To investigate the presence of an embryonic hourglass in plants, we combine this transcriptome information with two different measures of evolutionary distance: evolutionary age and sequence divergence. We compute two different transcriptome indices for each gene, the transcriptome age index (TAI)⁷ based on evolutionary age, and the transcriptome divergence index (TDI) based on sequence divergence. We investigate the profiles of these two transcriptome indices across the seven sampled embryo stages, and ask if and to what degree they show an hourglass pattern similar to that found for zebrafish⁷ or *Drosophila*^{7,8}.

For calculating the TAI, we assign an evolutionary age to each gene in the *A. thaliana* genome by sorting each gene into its phylostratum, defined as the most distant phylogenetic node containing at least one species with a detectable homologue (Methods, Supplementary Fig. 1, Supplementary Tables 1 and 2). The resulting phylostratigraphic map¹² contains 13 phylostrata, PS1–PS13 (Fig. 1a). PS1 includes the evolutionarily oldest genes with homologous sequences in prokaryotes, and PS13 includes the evolutionarily youngest genes with no homologue in any other species.

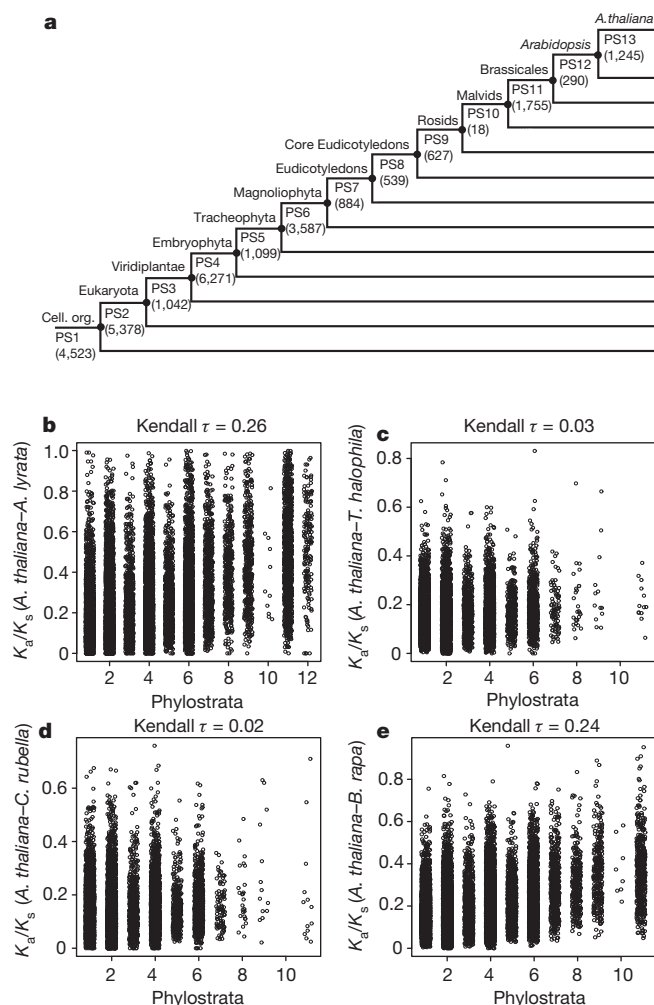


Figure 1 | Evolutionary age and sequence divergence of *A. thaliana* genes. **a**, Phylostratigraphic map of *A. thaliana*. Numbers in parenthesis denote the number of genes per phylostratum (PS1–PS13). Cell. org., cellular organisms described by PS1. **b–e**, Scatter plots of phylostratum versus K_a/K_s ratios over all genes. K_a/K_s ratios are derived from orthologous genes between *A. thaliana* and **b**, *A. lyrata*, **c**, *T. halophila*, **d**, *C. rubella* and **e**, *B. rapa*. Kendall τ values denote the Kendall rank correlation coefficients measuring the association between both parameters. See Methods for details.

¹Leibniz Institute of Plant Biochemistry, Independent Junior Research Group, Department of Molecular Signal Processing, Weinberg 3, 06120 Halle (Saale), Germany. ²Institute of Computer Science, Martin Luther University Halle–Wittenberg, 06120 Halle (Saale), Germany. ³UFZ Helmholtz Centre for Environmental Research, Department of Soil Ecology, 06120 Halle (Saale), Germany.

For calculating the TDI, we determine the sequence divergence between *A. thaliana* and its sister species *Arabidopsis lyrata* or any one of the closely related Brassicaceae, *Brassica rapa*, *Capsella rubella* and *Thellungiella halophila*, by computing the K_a/K_s ratio (Supplementary Table 3). Here K_a is the number of non-synonymous substitutions per non-synonymous site and K_s is the number of synonymous substitutions per synonymous site for each orthologous gene pair. The K_a/K_s ratio is an indicator of selective pressure within protein coding regions and, thus, reflects natural selection, one of the major forces driving molecular evolution. Interestingly, evolutionary age and sequence divergence as quantified above show only weak correlations (Kendall's rank correlation coefficient ranging from 0.02 to 0.26; Fig. 1b–e), indicating that both measures of evolutionary distance can be regarded as complementary (Supplementary Note).

In combination with transcript information, the TAI quantifies the mean evolutionary age of a transcriptome, where the evolutionary age (phylostratum) of each gene is weighted by its expression level⁷. Analogously, we define the TDI as the mean sequence divergence of a transcriptome, where the sequence divergence (K_a/K_s) of each gene is weighted by its expression level (Methods).

Figure 2 and Supplementary Fig. 2 show the TAI and TDI profiles across the seven sampled embryo stages of *A. thaliana*. We find that transcriptomes of early plant embryonic stages such as zygote and quadrant are evolutionarily young (high TAI), transcriptomes of the mid-embryogenic phase ranging from the globular to the torpedo stage are older (low TAI), and transcriptomes of later stages of embryogenesis are younger again (Fig. 2a). Qualitatively, this TAI profile strikingly resembles the molecular hourglass pattern discovered for zebrafish and *Drosophila*⁷. Likewise, we find that transcriptomes of early stages are divergent (high TDI), transcriptomes of the mid-embryogenic phase are more conserved (low TDI), and transcriptomes of later stages of embryogenesis are more divergent again (Fig. 2b). Remarkably, the TDI profile qualitatively resembles the molecular hourglass pattern of the gene expression divergence profile discovered for *Drosophila*⁸ and recently also *Caenorhabditis*¹³.

Comparing both profiles, we make two observations. First, each of the profiles shows an hourglass pattern, where the TAI reflects long-term evolutionary changes covering 4 billion years since the origin of life, and the TDI reflects short-term evolutionary changes covering

roughly 5–16 million years since the divergence of *A. thaliana* and the other four Brassicaceae^{14–17} (Supplementary Note). Second, both profiles point to the torpedo stage as the predicted phylotypic stage, representing simultaneously the stage with the oldest as well as the most conserved/least divergent transcriptome. An independent, but comparable transcriptome dataset^{18,19} from *A. thaliana* (Supplementary Fig. 3), which likewise covers embryogenesis from early phases to the mature embryo, confirms the hourglass pattern for both indices (Supplementary Fig. 4). Together, these observations suggest the possibility of convergent evolution of a molecular embryonic hourglass in animals and *A. thaliana*, and make it tempting to conjecture its universal presence across animal and plant kingdoms.

Given that developmental processes during plant and animal embryogenesis can be very different from the zygote stage on²⁰, and that the embryonic hourglass must have evolved independently in plants and animals, we wish to understand how the torpedo stage as the *bona fide* phylotypic stage in *A. thaliana* relates to animal phylotypic stages.

Across different animal taxa, the phylotypic stage was defined as the stage at which all major body parts are represented at their final positions as undifferentiated cell condensations²¹. In relation to this ontogenic progression in animals, the mid-embryogenic transition from the globular to the heart stage may conceptually serve as the corresponding stage in flowering plants. Here, polar axes are established and shoot and root apical meristems are initiated²². Hence, the ensuing torpedo stage at the transition from mid- to late-embryogenesis marks an ontogenic progression that seems more advanced than the phylotypic stage known from animals. Considering that morphological diversity and many important organs in flowering plants develop post-embryogenically, it is possible that the phylotypic stage may be shifted towards the transition from mid- to late-embryogenesis compared to animals.

Furthermore, the torpedo stage roughly marks the transition from morphogenesis to the maturation phase. Morphogenesis involves the establishment of the embryo's body plan, whereas maturation involves cell expansion and accumulation of storage macromolecules to prepare for desiccation, germination and early seedling growth²³. Whereas all land plants/embryophytes (all species from PS4 on) including lower land plants pass through a morphogenesis phase, only the embryogenesis of higher land plants concludes with a maturation phase. Completely different signalling cascades are involved in both phases. One set is switched off and the other one is initiated. Because torpedo stage embryos are in the transition between these different developmental programs, it is conceivable that transcriptional programs are likewise reduced to conserved and evolutionary ancient processes that are reflected by the neck of the hourglass (Fig. 2).

Encouraged by these findings, we seek to understand how the molecular hourglass pattern of the TAI profile is determined. Two simple scenarios that would result in a decrease of TAI values include up-regulation of old genes, or down-regulation of young genes during mid-embryogenesis. To distinguish between both scenarios, we compute the relative expression levels of genes from phylostrata containing pre-embryogenesis species (PS1–PS3) versus post-embryogenesis phylostrata (land plants/embryophytes from PS4–PS13, representing plant species that pass through embryogenesis). Whereas expression levels of old genes vary only marginally across embryo stages, young genes are down-regulated towards the torpedo stage, and the ratio of the relative expression levels of old and young genes is maximized in the torpedo stage (Fig. 3a, Supplementary Fig. 5). Next, we divide the genes along the median of the K_a/K_s ratios over all genes and perform an analogous analysis for conserved (below median) versus divergent (above median) genes. Interestingly, we find a similar pattern, with divergent genes being more down-regulated towards the torpedo stage than conserved genes (Fig. 3b; Supplementary Figs 6–9). These results are confirmed by the independent dataset^{18,19} (Supplementary Figs 10–14).

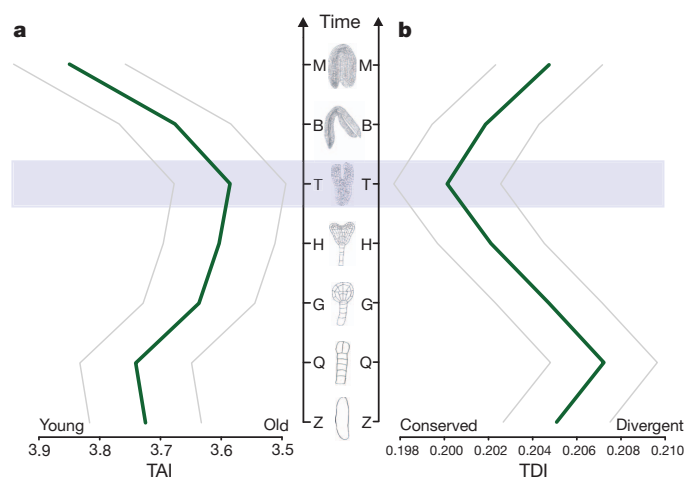


Figure 2 | Transcriptome indices across *A. thaliana* embryogenesis. **a**, The transcriptome age index (TAI) profile. **b**, The transcriptome divergence index (TDI) profile. Embryo stages: Z, zygote; Q, quadrant; G, globular; H, heart; T, torpedo; B, bent cotyledon; M, mature. Representative drawings (not on the same scale) are given for each sampled embryo stage. The blue shaded area marks the predicted phylotypic stage. The grey lines represent the standard error estimated by bootstrap analysis. The overall patterns of the TAI and TDI profiles are highly significant, as measured by permutation tests ($P_{TAI} = 2.55 \times 10^{-17}$; $P_{TDI} = 2.15 \times 10^{-33}$). See Methods for details.

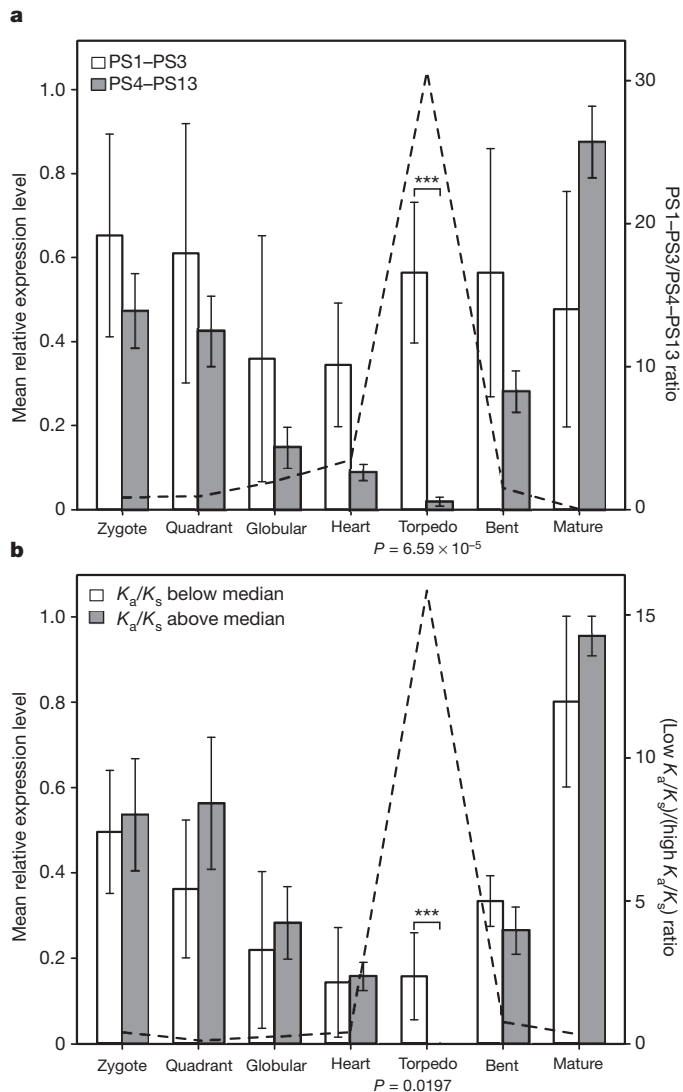


Figure 3 | Relative expression levels over embryo stages. **a**, Left axis, mean relative expression levels of genes in PS1–PS3 (open bars) and PS4–PS13 (shaded bars); right axis, ratio of mean relative expression levels between PS1–PS3 and PS4–PS13, data points connected by dashed line. **b**, Analogously to **a**, genes were divided along the median of the K_a/K_s ratios over all genes. Open and shaded bars show K_a/K_s values respectively below and above the median; data points connected by dashed line show the ratio of low to high K_a/K_s values. Error bars, standard error. Asterisks denote significant differences between PS1–PS3 and PS4–PS13 values (**a**) and conserved (below median) versus divergent (above median) genes at the torpedo stage (**b**); * $P < 0.05$; *** $P < 0.0005$.

Hence, the embryonic hourglass in *A. thaliana* seems to be coordinated by the quantitative down-regulation of young/divergent genes or, qualitatively, by the expression of fewer young/divergent genes towards the torpedo stage. This is in notable agreement with observations from the animal kingdom^{7,8,24,25}. As only a fraction of these down-regulated young genes in *A. thaliana* display an hourglass shaped expression profile across the sampled embryo stages themselves, the hourglass pattern is most probably caused by different sets of young genes. One set is involved in morphogenesis (up-regulated before the torpedo stage and down-regulated thereafter) and one set is involved in maturation (down-regulated before the torpedo stage and up-regulated thereafter; Supplementary Figs 15 and 16). In addition, we find that significantly enriched gene ontology terms among young plant genes down-regulated in the torpedo stage compared to early- or late embryogenesis describe signalling processes, such as responses to endogenous stimuli and hormones (Supplementary

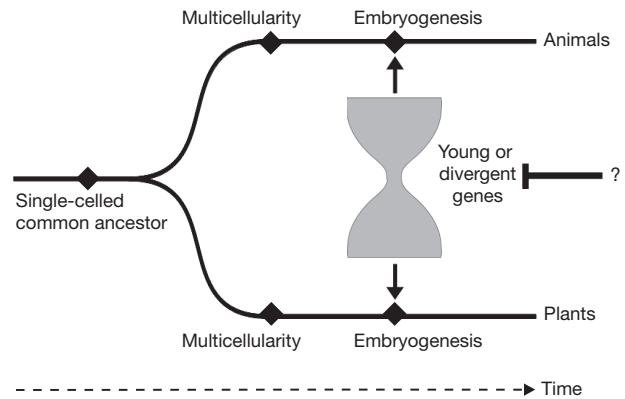


Figure 4 | Convergent evolution of a molecular hourglass in animal and plant embryogenesis. Originating from a single-celled common ancestor, animal and plant lineages evolved both multicellularity and embryogenesis independently. For the coordinated progression of the organisms through embryogenesis, the transcriptomes have to follow an hourglass pattern with maximally ancient and conserved transcriptomes during the phylotypic stage.

Tables 5 and 6). This indicates that signalling processes controlling transcription of relatively recently evolved genes are down-regulated during the predicted phylotypic stage of *A. thaliana* embryogenesis.

Using a phylotranscriptomic approach based on two complementary measures of evolutionary distance and two independent datasets, we have observed a molecular embryonic hourglass in plants, which seems to be predominantly caused by down-regulation of young and divergent genes towards the torpedo stage (Fig. 4). This observation is surprising for two reasons. First, morphological diversity during embryogenesis of flowering plants is negligible, so the increase of both transcriptome indices in late embryogenesis precedes the morphological differences established only during post-embryonic development. Second, convergent evolution of a molecular hourglass pattern in animals and plants suggests operation of a fundamental developmental profile controlling the expression of evolutionarily young or rapidly evolving genes across kingdoms. We speculate that such a mechanism may be required for enabling spatio-temporal organization and differentiation of complex multicellular life.

METHODS SUMMARY

Following ref. 7, 1,459 species with completely sequenced genomes (Supplementary Table 1) were sorted into 13 phylostrata (Supplementary Fig. 1), and amino acid sequences of *A. thaliana* were compared with the amino acid sequences of these species using Blast. If no Blast hit was identified, the corresponding gene of *A. thaliana* was assigned to PS13; otherwise, it was assigned to the phylogenetically most distant phylostratum containing at least one Blast hit. This procedure resulted in the phylostratigraphic map shown in Fig. 1a.

Orthologous gene pairs of *A. thaliana* and *A. lyrata*, *T. halophila*, *C. rubella*, or *B. rapa* were determined with the method of best hits using Blastp. Amino acid sequence alignments of each pair generated with MAFFT²⁶ (L-INS-i option) were used for codon alignments generated with PAL2NAL²⁷ to compute sequence divergence levels (K_a/K_s) with GEMESTATOR²⁸. Gene pairs with $K_a < 0.5$, $K_s < 5$ and K_a/K_s ratios < 2 were retained.

Introduced in ref. 7, the TAI of developmental stage s is the weighted mean of the phylostratum ps_i of gene i weighted by the expression level e_{is} of gene i at developmental stage s

$$TAI_s = \frac{\sum_{i=1}^n ps_i e_{is}}{\sum_{i=1}^n e_{is}} \quad (1)$$

where n is the total number of genes analysed. Low PS values correspond to evolutionarily old genes, so low TAI values correspond to evolutionarily old transcriptomes. Likewise, high PS values correspond to evolutionarily young genes, so high TAI values correspond to evolutionarily young transcriptomes.

Analogously, we introduce the TDI of developmental stage s by replacing ps_i in equation (1) by the K_a/K_s ratio of gene i :

$$\text{TDI}_s = \frac{\sum_{i=1}^n \left(\frac{K_{ai}}{K_{si}} \right) e_{is}}{\sum_{i=1}^n e_{is}} \quad (2)$$

Here, low/high K_a/K_s ratios correspond to conserved/divergent genes, so low/high TDI values correspond to conserved/divergent transcriptomes.

Full Methods and any associated references are available in the online version of the paper.

Received 26 January; accepted 3 July 2012.

Published online 5 September 2012.

- Meyerowitz, E. M. Plants compared to animals: the broadest comparative study of development. *Science* **295**, 1482–1485 (2002).
- Meckel, J. F. *Beyträge zur vergleichenden Anatomie* (Reclam, Leipzig, 1811).
- von Baer, K. E. *Über Entwicklungsgeschichte der Thiere: Beobachtung und Reflexion* (Gebrüder Bornträger, Königsberg, 1828).
- Duboule, D. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev., Suppl.* **1994**, 135–142 (1994).
- Raff, R. A. *The Shape of Life: Genes, Development and the Evolution of Animal Form* (Univ. Chicago Press, 1996).
- Sander, K. in *Development and Evolution* (eds Goodwin, B. C., Holder, N. & Wylie, C. C.) 137–159 (Cambridge Univ. Press, 1983).
- Domazet-Lošo, T. & Tautz, D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815–818 (2010).
- Kalinka, A. T. *et al.* Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–814 (2010).
- De Smet, I., Lau, S., Mayer, U. & Jürgens, G. Embryogenesis — the humble beginnings of plant life. *Plant J.* **61**, 959–970 (2010).
- Peris, C. I., Rademacher, E. H. & Weijers, D. Green beginnings — pattern formation in the early plant embryo. *Curr. Top. Dev. Biol.* **91**, 1–27 (2010).
- Xiang, D. *et al.* Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in *Arabidopsis*. *Plant Physiol.* **156**, 346–356 (2011).
- Domazet-Lošo, T., Brajković, J. & Tautz, D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* **23**, 533–539 (2007).
- Levin, M., Hashimshony, T., Wagner, F. & Yanai, I. Developmental milestones punctuate gene expression in the *Caenorhabditis* embryo. *Dev. Cell* **22**, 1101–1108 (2012).
- Koch, M. A., Haubold, B. & Mitchell-Olds, T. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol. Biol. Evol.* **17**, 1483–1498 (2000).
- Arakaki, M. *et al.* Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc. Natl Acad. Sci. USA* **108**, 8379–8384 (2011).
- Koch, M. A. & Kiefer, M. Genome evolution among cruciferous plants: a lecture from the comparison of the genetic maps of three diploid species — *Capsella rubella*, *Arabidopsis lyrata* subsp. *petraea*, and *A. thaliana*. *Am. J. Bot.* **92**, 761–767 (2005).
- Oh, D.-H. *et al.* Genome structures and halophyte-specific gene expression of the extremophile *Thellungiella parvula* in comparison with *Thellungiella salsuginea* (*Thellungiella halophila*) and *Arabidopsis*. *Plant Physiol.* **154**, 1040–1052 (2010).
- Zuber, H. *et al.* The seed composition of *Arabidopsis* mutants for the group 3 sulfate transporters indicates a role in sulfate translocation within developing seeds. *Plant Physiol.* **154**, 913–926 (2010).
- Le, B. H. *et al.* Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proc. Natl Acad. Sci. USA* **107**, 8063–8070 (2010).
- Nodine, M. D. & Bartel, D. P. Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* **482**, 94–97 (2012).
- Slack, J. M., Holland, P. W. & Graham, C. F. The zootype and the phylotypic stage. *Nature* **361**, 490–492 (1993).
- Lau, S., Slane, D., Herud, O., Kong, J. & Jürgens, G. Early embryogenesis in flowering plants: setting up the basic body pattern. *Annu. Rev. Plant Biol.* **63**, 483–506 (2012).
- Park, S. & Harada, J. J. *Arabidopsis* embryogenesis. *Methods Mol. Biol.* **427**, 3–16 (2008).
- Irie, N. & Kuratani, S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nature Commun.* **2**, 248 (2011).
- Irie, N. & Sehara-Fujisawa, A. The vertebrate phylotypic stage and an early bilaterian-related stage in mouse embryogenesis defined by genomic information. *BMC Biol.* **5**, 1 (2007).
- Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
- Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).
- Thornton, K. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* **19**, 2325–2327 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to S. Abel, L. I. A. Calderón Villalobos, C. Delker, T. Greb, D. Grubb, J. J. Harada, D. Jackson, R. Paxton, A. Soro and C. Wasternack for discussions, to S. Neumann for support with the IPB-cluster, and the 'Exzellenznetzwerk für Biowissenschaften' of the Federal State of Sachsen-Anhalt, Germany, for financial support of M.Q.

Author Contributions M.Q. conceived the study. M.Q. and I.G. supervised the project. H.-G.D., A.G., K.K.U. and M.B. analysed the data. M.Q., H.-G.D., A.G., K.K.U., M.B. and I.G. interpreted the results. M.Q. and I.G. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.Q. (mquint@ipb-halle.de).

METHODS

Phylostratigraphic procedure. A full account of the procedure of constructing a phylostratigraphic map has been presented previously^{7,12}. Adapted to *A. thaliana*, the following procedure was used in this study. The phylogeny of *A. thaliana* has been assigned according to the NCBI taxonomy database. For each of the 13 phylostrata shown in Supplementary Fig. 1, all 6,617,032 amino acid sequences of all 1,459 species with completely sequenced genomes were extracted from Phytozome²⁹ and the databases listed in Supplementary Table 1. A database of these sequences was generated, and each of the 27,258 amino acid sequences of *A. thaliana* (TAIR9) with a minimum length of 30 amino acids was blasted against this database using blastp (BLAST version 2.2.21) with an E-value cut-off of 10^{-5} . If no blast hit was identified, the corresponding gene of *A. thaliana* was assigned to phylostratum 13 (PS13). Otherwise, the corresponding gene of *A. thaliana* was assigned to the phylogenetically most distant (oldest) phylostratum containing at least one species with at least one blast hit (Supplementary Table 2). This procedure resulted in the phylostratigraphic map shown in Fig. 1a. To investigate the dependence of the results on the blastp stringency, the entire procedure was repeated for varying E-value cut-offs of blastp ranging from 10^{-1} to 10^{-20} , resulting in the TAI profiles shown in Supplementary Fig. 17.

K_a/K_s ratios. Orthologous gene pairs of *A. thaliana* and *A. lyrata* or the closely related Brassicaceae *T. halophila*, *C. rubella* or *B. rapa* were determined with the method of best hits using blastp. Amino acid sequence alignments of each pair generated with MAFFT²⁶ (L-INS-i option) were used for codon alignments generated with PAL2NAL²⁷ to compute sequence divergence levels (K_a/K_s) with GEMESTATOR²⁸. Gene pairs with $K_a < 0.5$, $K_s < 5$ and K_a/K_s ratios < 2 were retained (Supplementary Tables 3 and 4).

TAI and TDI. Expression levels for the first dataset were extracted from ref. 11, and after outlier detection and ID mapping, 25,158 genes represented on the microarrays were included in subsequent analyses. The TAI and the TDI are weighted means of evolutionary age and sequence divergence, respectively, and defined as follows. Introduced in ref. 7, the transcriptome age index TAI_s of developmental stage s (s = zygote, quadrant, globular, heart, torpedo, bent cotyledon, or mature) is the weighted mean of the evolutionary age (phylostratum) ps_i of gene i weighted by the expression level e_{is} of gene i at developmental stage s ,

$$TAI_s = \frac{\sum_{i=1}^n ps_i e_{is}}{\sum_{i=1}^n e_{is}}$$

where n is the total number of genes analysed. Low PS values correspond to evolutionarily old genes, so low TAI values correspond to evolutionarily old transcriptomes. Likewise, high PS values correspond to evolutionarily young genes, so high TAI values correspond to evolutionarily young transcriptomes.

Analogously, we introduce the transcriptome divergence index TDI_s of developmental stage s simply by replacing ps_i in the above equation by the K_a/K_s ratio of gene i ,

$$TDI_s = \frac{\sum_{i=1}^n \left(\frac{K_a}{K_s} \right) e_{is}}{\sum_{i=1}^n e_{is}}$$

Hence, low/high K_a/K_s ratios correspond to conserved/divergent genes, so low/high TDI values correspond to conserved/divergent transcriptomes.

The same procedure was repeated for the second independent dataset covering the embryo proper of *A. thaliana* embryo stages pre-globular, globular, heart, linear cotyledon/torpedo, and mature^{18,19} (GEO accession number GSE12404). We normalized this dataset using the GCRMA package (version 2.0) from the Bioconductor project with default parameter settings³⁰. For each probe set we computed the stage-wise arithmetic mean of the replicates to get representative expression values for each stage. Here, 20,031 genes represented on the microarrays were included in the analyses.

Statistical significance of TAI and TDI profiles. To determine the statistical significance of the TAI and TDI profiles, the following permutation test was performed. The variance V_{TAI} of the seven values of TAI_s (for s = zygote, quadrant, ..., mature) was computed as test statistic. For determining the null distribution of V_{TAI} , all PS values within each developmental stage s were randomly permuted, seven surrogate values of TAI_s were computed from this permuted dataset, and a surrogate value of V_{TAI} was computed from these seven surrogate values of TAI_s .

This procedure was repeated 1,000 times, yielding a histogram of 1,000 values of V_{TAI} , which can be approximated by a gamma distribution. The two parameters of the gamma distribution were estimated by the method of moments, the fitted gamma distribution was considered the null distribution of V_{TAI} , and the P -value of the observed value of V_{TAI} was computed from this null distribution.

The same procedure was repeated for the seven values of TDI_s , yielding a P -value of the TDI profile. Likewise, the second dataset^{18,19} was analysed accordingly.

Relative expression levels for phylostrata. Relative expression levels were computed as described previously⁷. In brief, the mean expression level e_{js} of phylostratum j and developmental stage s was computed for each j and s as the arithmetic mean of expression levels e_{is} of all genes i belonging to phylostratum j . The mean expression levels e_{js} were linearly transformed to the interval $[0,1]$ according to

$$f_{js} = \frac{e_{js} - e_{jmin}}{e_{jmax} - e_{jmin}}$$

where e_{jmin}/e_{jmax} is the minimum/maximum mean expression level of phylostratum j over the seven developmental stages s . This linear transformation corresponds to a shift by e_{jmin} and a subsequent shrinkage by $e_{jmax} - e_{jmin}$. As a result, the relative expression level f_{js} of developmental stage s with minimum e_{js} is 0, the relative expression level f_{js} of the developmental stage s with maximum e_{js} is 1, and the relative expression levels f_{js} of all other stages s range between 0 and 1, accordingly.

Next, relative expression levels were grouped into two PS classes, where the first PS class consists of relative expression levels of genes belonging to the three oldest phylostrata PS1–PS3, and where the second PS class consists of relative expression levels of genes belonging to the younger phylostrata PS4–PS13. This grouping was chosen to distinguish phylostrata of plants that pass through embryogenesis (PS4–PS13) from the remaining phylostrata (PS1–PS3), in which the vast majority of species did not evolve embryogenesis.

For each developmental stage s and each PS class, the mean value and standard error of the relative expression levels were computed. In addition, the ratio (fold-change) of the two relative expression levels was computed for each developmental stage s , and Welch's two-sample t -test was performed.

Relative expression levels for K_a/K_s quantiles. In contrast to PS values, which are discrete, K_a/K_s ratios are continuous. For computing relative expression levels of genes belonging to different K_a/K_s groups, continuous K_a/K_s ratios were grouped into deciles (10% quantiles). Relative expression levels of these ten K_a/K_s groups were computed in analogy to the computation of relative expression levels of the 13 phylostrata.

Likewise, relative expression levels were grouped into two K_a/K_s classes, where the first K_a/K_s class consists of relative expression levels of genes belonging to the first five K_a/K_s groups (K_a/K_s ratios below median, conserved genes), and where the second K_a/K_s class consists of relative expression levels of genes belonging to the remaining five K_a/K_s groups (K_a/K_s ratios above median, divergent genes). This grouping was chosen because the median is a natural choice, making both K_a/K_s classes equally large, and because the grouping of genes into different PS classes also resulted in two PS classes of roughly similar sizes (first dataset: 10,695 genes in PS1–PS3 and 14,463 genes in PS4–PS13; second dataset: 9,028 and 11,003 genes, respectively).

The computation of mean values, standard errors, fold-changes and P -values of Welch's two-sample t -test were performed as described in the previous section.

To investigate the dependence of the results on the grouping into two K_a/K_s classes, the entire analysis was repeated for the following six pairs of K_a/K_s classes: two deciles/eight deciles, three deciles/seven deciles, ..., seven deciles/three deciles. The six resulting plots of means, standard errors, fold-changes and P -values are presented in Supplementary Figs 6–9, 11–14.

Gene ontology analysis. Enrichment analysis of gene ontology terms was performed on genes from PS4–PS13 that were down-regulated at least two-fold in the torpedo stage compared to at least one of the developmental stages zygote/quadrant/globular/heart/bent cotyledon/mature (first dataset), or pre-globular/globular/heart/mature (second dataset). Gene ontology term enrichment was analysed using AmiGO³¹ (filter options: TAIR and *A. thaliana*, P -value = 1×10^{-2}).

29. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186 (2012).
30. Wu, Z. *et al.* A model-based background adjustment for oligonucleotide expression arrays. *J. Am. Stat. Assoc.* **99**, 909–917 (2004).
31. Carbon, S. *et al.* AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288–289 (2009).

Pregnancy imprints regulatory memory that sustains anergy to fetal antigen

Jared H. Rowe¹, James M. Ertelt^{1†}, Lijun Xin^{1†} & Sing Sing Way^{1†}

Pregnancy is an intricately orchestrated process where immune effector cells with fetal specificity are selectively silenced. This requires the sustained expansion of immune-suppressive maternal FOXP3⁺ regulatory T cells (T_{reg} cells), because even transient partial ablation triggers fetal-specific effector T-cell activation and pregnancy loss^{1,2}. In turn, many idiopathic pregnancy complications proposed to originate from disrupted fetal tolerance are associated with blunted maternal T_{reg} expansion^{3–5}. Importantly, however, the antigen specificity and cellular origin of maternal T_{reg} cells that accumulate during gestation remain incompletely defined. Here we show that pregnancy selectively stimulates the accumulation of maternal FOXP3⁺ CD4 cells with fetal specificity using tetramer-based enrichment that allows the identification of rare endogenous T cells⁶. Interestingly, after delivery, fetal-specific T_{reg} cells persist at elevated levels, maintain tolerance to pre-existing fetal antigen, and rapidly re-accumulate during subsequent pregnancy. The accelerated expansion of T_{reg} cells during secondary pregnancy was driven almost exclusively by proliferation of fetal-specific FOXP3⁺ cells retained from prior pregnancy, whereas induced FOXP3 expression and proliferation of pre-existing FOXP3⁺ cells each contribute to T_{reg} expansion during primary pregnancy. Furthermore, fetal resorption in secondary compared with primary pregnancy becomes more resilient to partial maternal FOXP3⁺ cell ablation. Thus, pregnancy imprints FOXP3⁺ CD4 cells that sustain protective regulatory memory to fetal antigen. We anticipate that these findings will spark further investigation on maternal regulatory T-cell specificity that unlocks new strategies for improving pregnancy outcomes and novel approaches for therapeutically exploiting T_{reg} cell memory.

The accumulation of maternal T_{reg} cells during pregnancy parallels the need for expanded tolerance to encompass 'non-self' fetal antigens^{3–5,7,8}. However, one consequence of sustained FOXP3⁺ cell expansion is susceptibility to prenatal infection⁷. Given the increasingly recognized importance of T_{reg} specificity in regulating the fluid balance between immune activation that maintains host defence and immune suppression that prevents autoimmunity^{9–14}, we reasoned that establishing the specificity of maternal T_{reg} cells that expand during pregnancy could unravel ways to dissociate their beneficial and detrimental impacts. Furthermore, extending this analysis post-partum may allow the regulatory memory recently described for T_{reg} cells responsive to an induced self antigen to be investigated in a more physiological context¹⁵. To address these questions, we developed a mating strategy where the I-A^b 2W1S_{55–68} peptide (a variant of peptide residues 55–68 for the alpha chain of the mouse major histocompatibility complex (MHC) class II, I-E^d) becomes a surrogate fetal antigen using male mice (H-2^d; Balb/c or H-2^b C57BL/6 [B6]) engineered to co-express this peptide with β-actin to impregnate non-2W1S-expressing B6 females¹⁶. In turn, the high precursor frequency of CD4 cells with 2W1S_{55–68} specificity allows endogenous maternal T_{reg} cells to this surrogate fetal antigen to be identified using MHC class II tetramer enrichment⁶.

Using this approach, maternal CD4 cells with fetal-2W1S specificity were found to sharply upregulate CD44 expression, progressively accumulate throughout pregnancy, and persist at approximately ten-fold increased levels through day 100 post-partum compared with non-pregnant controls (Fig. 1a). Maternal 2W1S⁺ cell expansion was specific to mating with 2W1S-expressing mice because they did not accumulate in females impregnated by non-transgenic Balb/c males (Supplementary Fig. 1). Because seminal fluid also contains cells of paternal origin¹⁷, 2W1S⁺ cells in female mice rendered infertile with low-dose irradiation were also enumerated. We found that although mating without pregnancy stimulated modest 2W1S⁺ cell expansion and CD44 upregulation, the magnitude was markedly reduced compared with pregnant mice (Supplementary Fig. 1). Thus, maternal 2W1S⁺ CD4 cell expansion during pregnancy reflects an antigen-specific response to cells of fetal origin.

Given the essential requirement for T_{reg} cells in maintaining fetal tolerance^{2,7,18,19}, we investigated FOXP3 expression among maternal cells with fetal-2W1S specificity. Beginning mid-gestation, 2W1S⁺ compared with 2W1S[−] CD4 cells became enriched for FOXP3 expression in allogeneic (Fig. 1a, b), as well as syngeneic pregnancy (Supplementary Fig. 2). As pregnancy progressed, FOXP3 expression among 2W1S⁺ cells became progressively more pronounced, peaking at around 50% late gestation through to the first 48 h post-partum (embryonic day 18.5 (E18.5) to post-partum day 2 (PP2)) (Fig. 1a, b and Supplementary Fig. 3). Furthermore, 2W1S⁺FOXP3⁺ cells, and to a lesser extent 2W1S⁺FOXP3[−] cells, upregulated the proliferation marker Ki67 that paralleled expanding fetal tissue (Fig. 1c). Reciprocally after expulsion of the fetus (PP14 to PP100), Ki67 expression among 2W1S⁺FOXP3⁺ and 2W1S⁺FOXP3[−] cells became reduced (Fig. 1c). However, despite diminished Ki67 levels, FOXP3 expression among 2W1S⁺ cells was sustained at ~20% through day 100 post-partum (Fig. 1a, b). Accordingly, maternal T_{reg} cells with fetal specificity selectively accumulate during pregnancy and persist following parturition.

Interestingly, maternal T_{reg} cells with fetal-2W1S specificity also progressively downregulated Helios (also known as IKZF2) expression that dropped to its lowest level of ~40% Helios^{hi} by late gestation, whereas the few 2W1S⁺FOXP3⁺ cells in non-pregnant mice were uniformly Helios^{hi} (Fig. 1d). Comparatively, Helios expression among bulk maternal T_{reg} cells did not shift significantly. Although this discordance in Helios expression may suggest conversion of fetal-specific FOXP3[−] cells into FOXP3⁺ cells²⁰, the recent finding that some peripherally induced T_{reg} cells also express Helios led us to more definitively investigate the origin of maternal T_{reg} cells with fetal specificity²¹. In particular, we asked whether mating with 2W1S-expressing males can convert 2W1S⁺FOXP3[−] CD4 cells from *Foxp3*^{DTR/DTR} donors ablated of T_{reg} cells with diphtheria toxin²² (*Foxp3*^{DTR/DTR} T_{reg} cells carry the human diphtheria toxin receptor (DTR) fused to an internal ribosome entry site into the 3' untranslated region of *Foxp3*, rendering them susceptible to ablation with low-dose diphtheria toxin) into FOXP3⁺ cells after adoptive transfer into virgin *Foxp3*^{WT/WT} recipient mice. By mid-gestation, 2W1S⁺FOXP3⁺ among T_{reg}-ablated donor *Foxp3*^{DTR/DTR}

¹University of Minnesota School of Medicine, Departments of Pediatrics and Microbiology, Center for Infectious Disease and Microbiology Translational Research, Center for Immunology, Minneapolis, Minnesota 55455, USA. †Present addresses: Cincinnati Children's Hospital Medical Center, Division of Infectious Diseases, Cincinnati, Ohio 45229, USA (J.M.E., L.X. and S.S.W.).

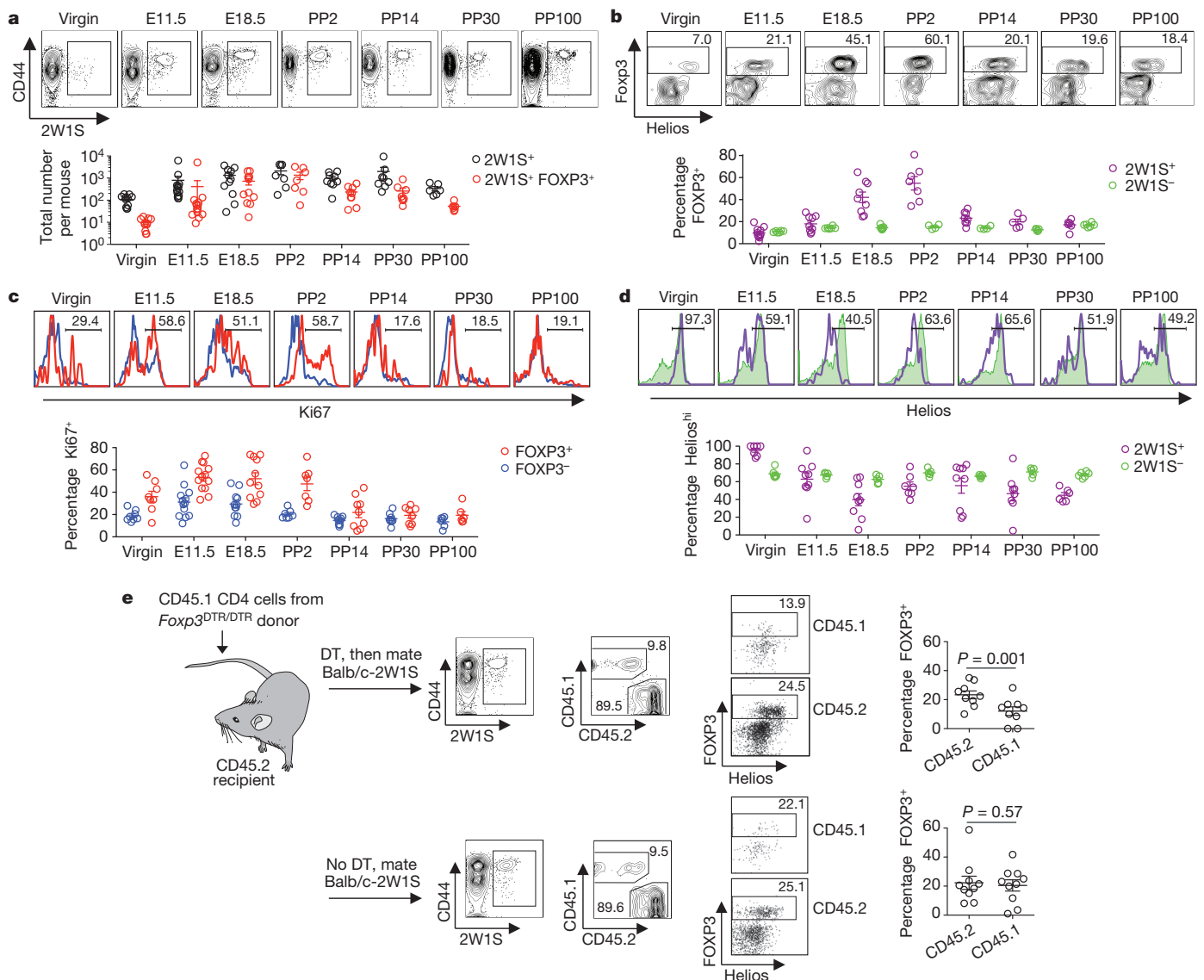


Figure 1 | Accumulation of maternal CD4 and FOXP3⁺ T_{reg} cells with fetal specificity during gestation. **a**, Total 2W1S⁺ or 2W1S⁺FOXP3⁺ CD4 cells in B6 females impregnated by Balb/c-2W1S males. **b**, Percentage FOXP3⁺ among 2W1S⁺ or 2W1S⁺FOXP3⁺ CD4 cells. **c**, Percentage Ki67⁺ among 2W1S⁺FOXP3⁺ or 2W1S⁺FOXP3⁺ CD4 cells. **d**, Percentage Helios^{hi} among

2W1S⁺FOXP3⁺ or 2W1S⁺FOXP3⁺ CD4 cells. **e**, Percentage FOXP3⁺ among *Foxp3*^{DTR/DTR} donor (CD45.1⁺) or *Foxp3*^{WT/WT} recipient (CD45.2⁺) 2W1S⁺ CD4 cells mid-gestation (E11.5) by Balb/c-2W1S males, with diphtheria toxin (DT) treatment (top) or no diphtheria toxin controls (bottom). Bars, means \pm one standard error.

cells were readily recovered, illustrating induction of maternal T_{reg} cells with fetal specificity (Fig. 1e). This conversion was pregnancy-specific and not due to incomplete donor T_{reg} ablation because FOXP3⁺ cells were undetectable among T_{reg}-ablated donor cells in unmated control mice (Supplementary Fig. 4). Importantly, however, FOXP3⁺ among T_{reg}-ablated donor CD4 cells was also consistently reduced (by \sim 50%) compared with either 2W1S⁺FOXP3⁺ donor cells in mice without diphtheria toxin treatment or among recipient CD4 cells not susceptible to diphtheria toxin (Fig. 1e). Thus, FOXP3 induction among FOXP3⁺ precursors and proliferation of pre-existing FOXP3⁺ cells each contribute to the accumulation of maternal T_{reg} cells with fetal specificity during primary pregnancy.

To further characterize maternal T_{reg} cells with specificity to pre-existing fetal antigen that persist post-partum, these cells were tracked during subsequent pregnancy. After secondary mating, maternal FOXP3⁺ cells with fetal-2W1S specificity accumulated with accelerated kinetics in an antigen-specific fashion (Fig. 2a and Supplementary Fig. 5). The more rapid expansion of maternal T_{reg} cells in separate groups of mice was recapitulated within the same mouse by measuring

2W1S⁺ T_{reg} accumulation among donor CD4 cells from post-partum mice (secondary expansion) adoptively transferred before mating with 2W1S-expressing males, compared with cells in virgin recipient mice (primary expansion) (Fig. 2b). By substituting CD4 cells from post-partum *Foxp3*^{DTR/DTR} mice for adoptive transfer and using diphtheria toxin to eliminate donor T_{reg} cells before mating, we also addressed whether the accelerated secondary expansion of maternal T_{reg} cells with fetal-2W1S specificity reflects more vigorous induction among FOXP3⁺ cells or proliferation of pre-existing FOXP3⁺ cells. We found that, in sharp contrast to primary pregnancy, the ablation of donor T_{reg} cells from post-partum *Foxp3*^{DTR/DTR} mice almost uniformly eliminated their expansion in subsequent pregnancy (Fig. 2c). Thus, recurrent pregnancy primes the accelerated accumulation of maternal FOXP3⁺ cells that expand from pre-existing T_{reg} cells retained from prior pregnancy.

Expanding this model, the responsiveness of maternal CD4 cells with fetal specificity was also investigated. We found 2W1S⁺ cells recovered from mice mid-gestation or post-partum each compared with non-pregnant controls did not produce appreciable IFN- γ *ex vivo*

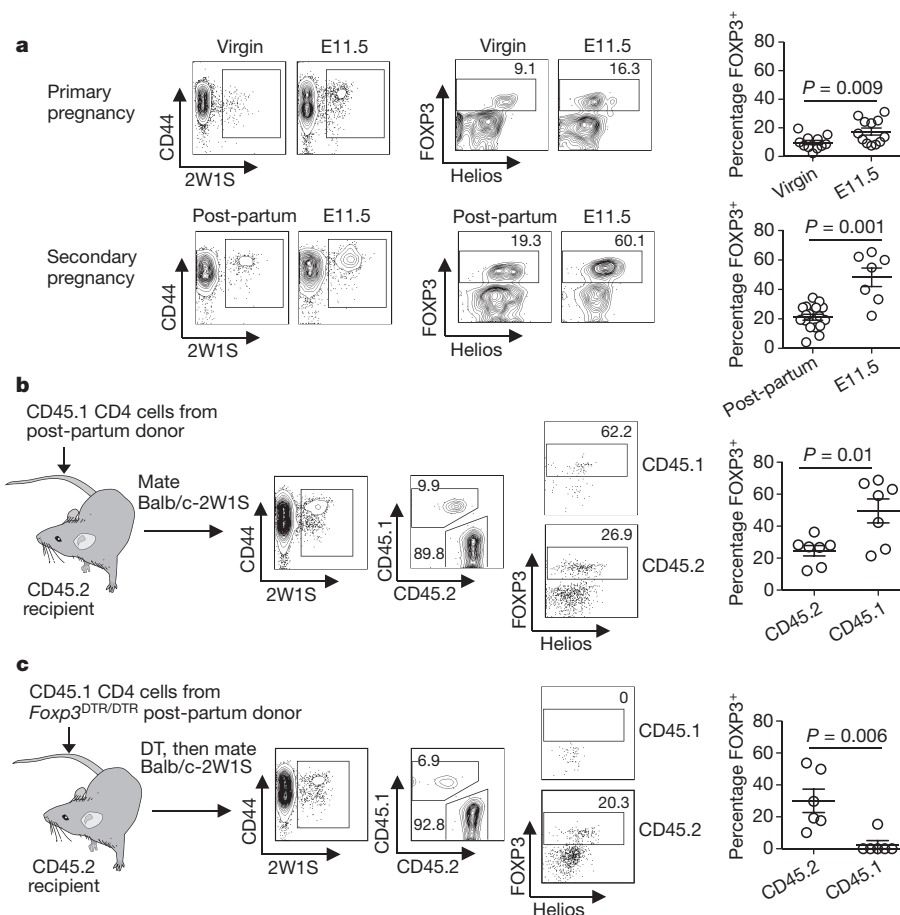


Figure 2 | Accelerated expansion of maternal T_{reg} cells with fetal specificity during secondary pregnancy. **a**, Percentage FOXP3⁺ among virgin (primary pregnancy) or post-partum (secondary pregnancy) females before mating or mid-gestation (E11.5) by Balb/c-2W1S males. **b**, Percentage FOXP3⁺ among

post-partum donor (CD45.1⁺) or naive recipient (CD45.2⁺) 2W1S⁺ CD4 cells mid-gestation by Balb/c-2W1S males. **c**, Percentage FOXP3⁺ among T_{reg}-ablated *Foxp3*^{DTR/DTR} post-partum donor (CD45.1⁺) or naive recipient (CD45.2⁺) 2W1S⁺ CD4 cells mid-gestation. Bars, means \pm one standard error.

following stimulation, consistent with previously described anergy of maternal cells with fetal specificity (Supplementary Fig. 6)²²³. Therefore, to more fully evaluate the responsiveness of maternal T cells with fetal-2W1S specificity, we measured their *in vivo* response to *Listeria monocytogenes* engineered to express the 2W1S_{55–68} peptide (Lm-2W1S) that potently stimulates T_H1-differentiation in other contexts^{24,25}. We found that 2W1S⁺ cells expand and upregulate T-bet (also known as TBX21) expression each in an antigen-specific fashion in both naive mice and mice impregnated by 2W1S-expressing males after Lm-2W1S inoculation, similar to other intracellular pathogens (Supplementary Fig. 7)²⁶. Interestingly, however, 2W1S⁺ cells in pregnant mice where 2W1S represents a surrogate fetal antigen produced only background levels of IFN- γ and other effector cytokines, with reciprocal accumulation of FOXP3⁺ cells (Fig. 3a and Supplementary Fig. 8). Comparatively, >15% of 2W1S⁺ cells in Lm-2W1S-inoculated virgin mice were IFN- γ ⁺ (Fig. 3a). This hypo-responsiveness was specific to fetal-2W1S stimulation, because 2W1S⁺ CD4 cells in mice impregnated with non-2W1S-expressing males produced IFN- γ levels comparable to non-pregnant controls (Fig. 3a and Supplementary Fig. 8). Given the sustained enrichment of fetal-specific T_{reg} cells after delivery (Fig. 1a, b), these studies were extended to investigate whether diminished IFN- γ production among maternal CD4 cells with specificity to pre-existing fetal antigen is similarly maintained. Remarkably, IFN- γ production remained anaemic in post-partum mice previously exposed to 2W1S as a fetal antigen, whereas post-partum mice without prior fetal-2W1S exposure produce IFN- γ comparable to non-pregnant controls (Fig. 3b). Accordingly, pregnancy imprints functional anergy for maternal CD4 cells with fetal specificity that is sustained post-partum.

To dissociate whether pregnancy-induced T-cell anergy was cell-intrinsic or imposed by features associated with the post-partum environment, we measured IFN- γ production by CD4 cells from post-partum or virgin mice after adoptive transfer into naive recipient mice. We found IFN- γ production by donor post-partum and each group of naive (donor and recipient) 2W1S⁺ CD4 cells were similar, and notably increased compared with 2W1S⁺ cells in un-manipulated post-partum mice following Lm-2W1S inoculation (Fig. 3b, c). Thus, anergy among maternal CD4 cells with specificity to pre-existing fetal antigen is not cell-intrinsic, but maintained by the post-partum environment.

In complementary studies we addressed the importance of maternal T_{reg} cells in sustaining anergy to cells with specificity to pre-existing fetal antigen by investigating the effect of replacing the entire T_{reg} compartment in post-partum mice previously exposed to fetal-2W1S with naive FOXP3⁺ cells from virgin mice. Consistent with recent studies using adoptively transferred *Foxp3*^{WT/WT} CD4 cells to refill the cellular compartment in *Foxp3*^{DTR/DTR} mice sustained on diphtheria toxin treatment², T_{reg} cells from naive mice efficiently reconstituted T_{reg}-ablated *Foxp3*^{DTR/DTR} post-partum mice (Fig. 4a). Using this approach, we found that replacing maternal FOXP3⁺ cells in post-partum mice with T_{reg} cells from naive mice restored IFN- γ production for 2W1S⁺ CD4 cells (Fig. 4b). Furthermore, whereas only rare 2W1S⁺FOXP3⁺ cells that were Helios^{hi} were found among post-partum mice reconstituted with naive T_{reg} cells, a significant proportion of 2W1S⁺ cells expanded in response to Lm-2W1S in intact post-partum mice remained FOXP3⁺ (~20%) and Helios^{lo} (~40%) (Fig. 4b). Thus, the muted expansion of naive FOXP3⁺ CD4 cells with *L. monocytogenes* infection is overcome by pregnancy-induced T_{reg} activation (Fig. 4b and

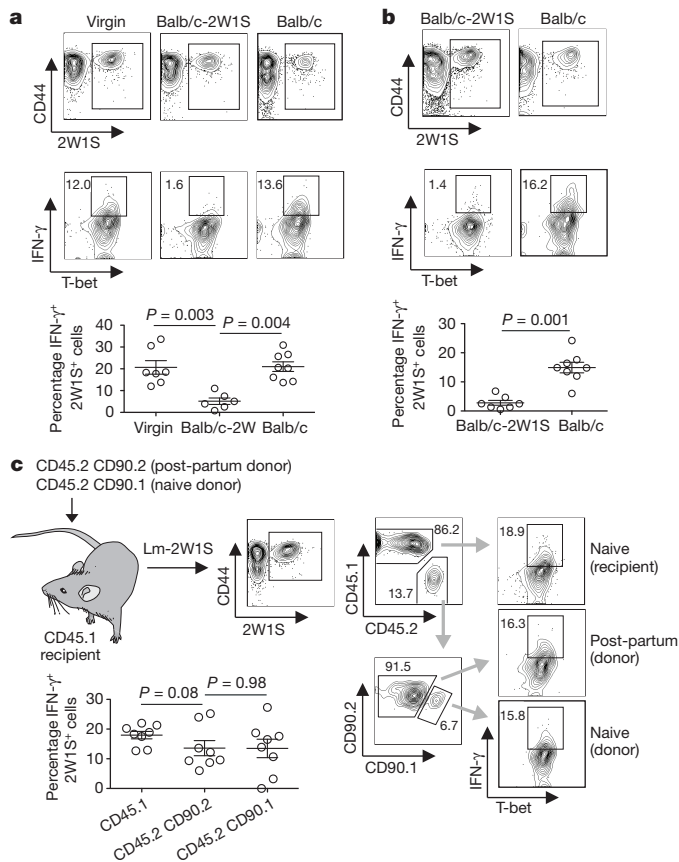


Figure 3 | The post-partum environment maintains energy for maternal CD4 cells with pre-existing fetal specificity. **a**, IFN- γ -producing 2W1S⁺ CD4 cells 5 days after Lm-2W1S inoculation in virgin or pregnant mice mid-gestation by Balb/c-2W1S or Balb/c males. **b**, IFN- γ -producing 2W1S⁺ CD4 cells 5 days after Lm-2W1S inoculation in post-partum mice previously impregnated by Balb/c-2W1S or Balb/c males. **c**, IFN- γ -producing post-partum donor (CD45.2⁺CD90.2⁺), naive donor (CD45.2⁺CD90.1⁺) or naive recipient (CD45.1⁺) CD4 cells 5 days after Lm-2W1S inoculation and stimulation with phorbol myristate acetate/ionomycin. Bars, means \pm one standard error.

Supplementary Fig. 8)²⁴. By extension, the restored responsiveness of post-partum 2W1S⁺ cells after adoptive transfer into T_{reg}-sufficient naive mice most likely represents dilution of co-transferred maternal T_{reg} cells with fetal-2W1S specificity (Fig. 3c).

Lastly, to establish how maternal T_{reg} cells with fetal specificity retained post-partum have an effect on subsequent pregnancy outcomes, the frequency of fetal resorption triggered by partial maternal FOXP3⁺ cell ablation using *Foxp3*^{DTR/WT} mice was compared between secondary and primary pregnancy². We found secondary pregnancy became significantly more resilient to partial T_{reg} ablation because fetal resorption was reduced by ~60% compared with primary pregnancy (Fig. 4c). In turn, fetal resorption in T_{reg}-sufficient *Foxp3*^{WT/WT} mice during secondary pregnancy was also significantly reduced from background levels compared with primary allogeneic pregnancy. Maternal T_{reg} cells were essential for these protective effects because wholesale FOXP3⁺ cell ablation using *Foxp3*^{DTR/DTR} mice triggered pervasive fetal resorption equally in secondary and primary allogeneic pregnancy (Fig. 4c). Importantly, fetal wastage with maternal T_{reg} ablation in this context was driven by antigen heterogeneity, and not poor maternal health, because the frequency of resorption was sharply reduced with T_{reg} ablation in mice bearing syngeneic pregnancy (Supplementary Fig. 9).

Together, these findings establish a model whereby pregnancy primes the selective accumulation and activation of maternal T_{reg} cells with fetal

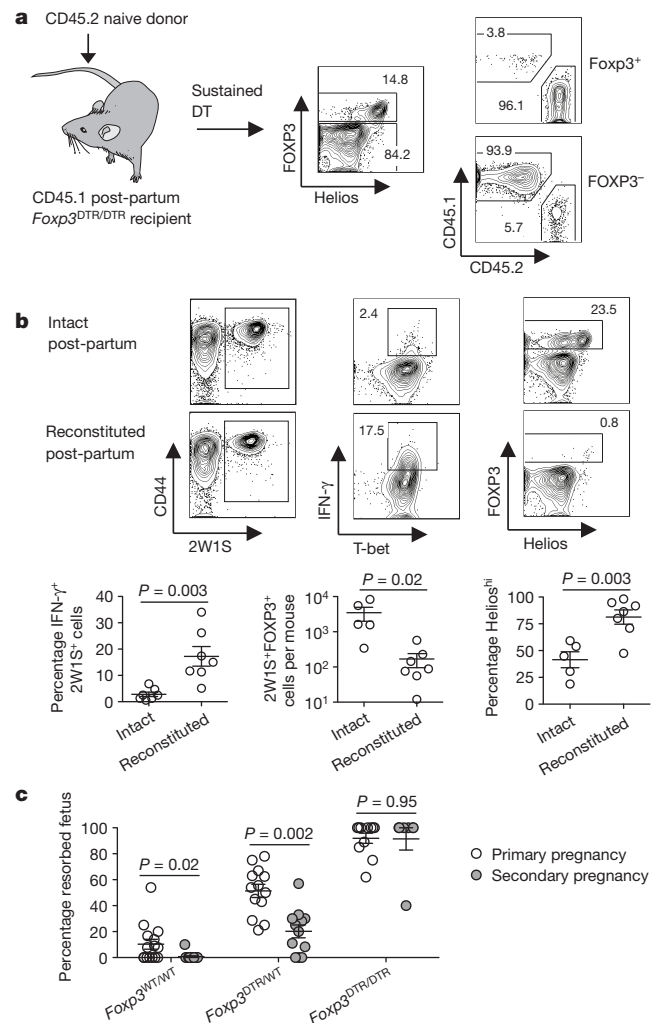


Figure 4 | Maternal post-partum T_{reg} cells mitigate IFN- γ responsiveness and mediate resiliency to fetal resorption in secondary pregnancy. **a**, Representative plots illustrating the majority (>96%) of FOXP3⁺ cells are derived from adoptively transferred CD4 in diphtheria-toxin-treated *Foxp3*^{DTR/DTR} mice. **b**, IFN- γ -producing 2W1S⁺ cells among CD45.1⁺CD45.2⁻ cells, accumulation of 2W1S⁺FOXP3⁺ cells, and Helios expression among 2W1S⁺FOXP3⁺ T_{reg} cells 5 days after Lm-2W1S inoculation. **c**, Percentage fetal resorption during primary (open) or secondary (shaded) allogeneic pregnancy for *Foxp3*^{WT/WT}, *Foxp3*^{DTR/WT} or *Foxp3*^{DTR/DTR} females 5 days after diphtheria toxin initiation beginning mid-gestation. Bars, means \pm one standard error.

specificity (Supplementary Fig. 10), and extend the role of antigen-experienced T_{reg} cells from primary into subsequent pregnancies^{2,7,18}. In this regard, whereas maternal T_{reg} cells have been described to expand up to twofold when examined in a non-antigen-specific fashion²⁻⁵, our results demonstrate that FOXP3⁺ cells with fetal specificity expand >100-fold through parturition (Fig. 1a and Supplementary Fig. 3). After delivery, maternal T_{reg} cells with fetal specificity are sustained at enriched levels, and are functionally distinct as they re-accumulate with accelerated kinetics and out-compete 'naive' T_{reg} cells during secondary pregnancy. Similar to discordant functional properties of naive and activated effector T cells²⁷, these results uncover the exciting possibility of exploiting antigen-specific 'memory' T_{reg} cells to dissociate detrimental and beneficial immune responses. Applied to human pregnancy, these data may explain why rates of pre-eclampsia, and other complications associated with disrupted fetal tolerance, are reduced in secondary compared with primary pregnancy²⁸. However, given the increased risk of pre-eclampsia in recurrent human pregnancy when the inter-pregnancy interval is extended, waning T_{reg} memory similar

to other CD4 subsets would not be unexpected^{25,29}. Therefore, establishing the durability of pregnancy-induced regulatory memory and characterizing whether this response can be sustained with boosting represent next steps with exceptional scientific importance.

METHODS SUMMARY

Mice. 2W1S-expressing and *Foxp3*^{DTR} mice have each been described^{16,22}. Mated females and the timing of pregnancy was determined by visualization of a copulation plug (E0.5). For infection, Lm-2W1S (10⁴ colony-forming units, see Methods)²⁴ was inoculated intravenously. Experiments were performed in accordance with University of Minnesota IACUC approved protocols.

Tetramer enrichment and flow cytometry. I-A^b 2W1S_{55–68} tetramer staining and enrichment have been described⁶. Lymphoid cells from the spleen and lymph nodes were enriched with 2W1S_{55–68} tetramer before surface, intracellular or intranuclear staining. For stimulation, cells were cultured with phorbol myristate acetate plus ionomycin (5 h) before staining.

Cell transfer and ablation. One mouse equivalent of purified CD4 cells was intravenously transferred into recipient mice 1 day before mating or infection. Donor T_{reg} cells from *Foxp3*^{DTR} mice were ablated in recipient *Foxp3*^{WT} mice using purified diphtheria toxin (two doses 8 h apart, 0.5 µg per dose). For ablation of endogenous T_{reg} cells, diphtheria toxin was administered daily (0.5 µg first dose, followed by 0.1 µg per dose daily thereafter) beginning mid-gestation or immediately following donor-cell transfer.

Statistics. Data were analysed using the unpaired (separate groups of mice) or paired (cell subsets within the same mouse) Student's *t* test.

Full Methods and any associated references are available in the online version of the paper.

Received 15 April; accepted 31 July 2012.

Published online 26 September 2012.

- Kahn, D. A. & Baltimore, D. Pregnancy induces a fetal antigen-specific maternal T regulatory cell response that contributes to tolerance. *Proc. Natl Acad. Sci. USA* **107**, 9299–9304 (2010).
- Rowe, J. H., Ertelt, J. M., Aguilera, M. N., Farrar, M. A. & Way, S. S. Foxp3⁺ regulatory T cell expansion required for sustaining pregnancy compromises host defense against prenatal bacterial pathogens. *Cell Host Microbe* **10**, 54–64 (2011).
- Prins, J. R. *et al.* Preeclampsia is associated with lower percentages of regulatory T cells in maternal blood. *Hypertens. Pregnancy* **28**, 300–311 (2009).
- Santner-Nanan, B. *et al.* Systemic increase in the ratio between Foxp3⁺ and IL-17-producing CD4⁺ T cells in healthy pregnancy but not in preeclampsia. *J. Immunol.* **183**, 7023–7030 (2009).
- Sasaki, Y. *et al.* Decidual and peripheral blood CD4⁺CD25⁺ regulatory T cells in early pregnancy subjects and spontaneous abortion cases. *Mol. Hum. Reprod.* **10**, 347–353 (2004).
- Moon, J. J. *et al.* Naive CD4⁺ T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* **27**, 203–213 (2007).
- Aluvihare, V. R., Kallikourdis, M. & Betz, A. G. Regulatory T cells mediate maternal tolerance to the fetus. *Nature Immunol.* **5**, 266–271 (2004).
- Andersen, K. G., Nissen, J. K. & Betz, A. G. Comparative genomics reveals key gain-of-function events in Foxp3 during regulatory T cell evolution. *Front. Immunol.* **3**, 113 (2012).
- Lathrop, S. K. *et al.* Peripheral education of the immune system by colonic commensal microbiota. *Nature* **478**, 250–254 (2011).
- Shafiani, S., Tucker-Heard, G., Kariyone, A., Takatsu, K. & Urdahl, K. B. Pathogen-specific regulatory T cells delay the arrival of effector T cells in the lung during early tuberculosis. *J. Exp. Med.* **207**, 1409–1420 (2010).
- Josefowicz, S. Z., Lu, L. F. & Rudensky, A. Y. Regulatory T cells: mechanisms of differentiation and function. *Annu. Rev. Immunol.* **30**, 531–564 (2012).
- Wing, K. & Sakaguchi, S. Regulatory T cells exert checks and balances on self tolerance and autoimmunity. *Nature Immunol.* **11**, 7–13 (2010).
- Hsieh, C. S., Lee, H. M. & Lio, C. W. Selection of regulatory T cells in the thymus. *Nat. Rev. Immunol.* **12**, 157–167 (2012).
- Suffia, I. J., Reckling, S. K., Piccirillo, C. A., Goldszmid, R. S. & Belkaid, Y. Infected site-restricted Foxp3⁺ natural regulatory T cells are specific for microbial antigens. *J. Exp. Med.* **203**, 777–788 (2006).
- Rosenblum, M. D. *et al.* Response to self antigen imprints regulatory memory in tissues. *Nature* **480**, 538–542 (2011).
- Moon, J. J. *et al.* Quantitative impact of thymic selection on Foxp3⁺ and Foxp3[−] subsets of self-peptide/MHC class II-specific CD4⁺ T cells. *Proc. Natl Acad. Sci. USA* **108**, 14602–14607 (2011).
- Robertson, S. A. Immune regulation of conception and embryo implantation—all about quality control? *J. Reprod. Immunol.* **85**, 51–57 (2010).
- Kallikourdis, M., Andersen, K. G., Welch, K. A. & Betz, A. G. Alloantigen-enhanced accumulation of CCR5⁺ 'effector' regulatory T cells in the gravid uterus. *Proc. Natl Acad. Sci. USA* **104**, 594–599 (2007).
- Munoz-Suano, A., Hamilton, A. B. & Betz, A. G. Gimme shelter: the immune system during pregnancy. *Immunol. Rev.* **241**, 20–38 (2011).
- Thornton, A. M. *et al.* Expression of Helios, an Ikaros transcription factor family member, differentiates thymic-derived from peripherally induced Foxp3⁺ T regulatory cells. *J. Immunol.* **184**, 3433–3441 (2010).
- Gottschalk, R. A., Corse, E. & Allison, J. P. Expression of Helios in peripherally induced Foxp3⁺ regulatory T cells. *J. Immunol.* **188**, 976–980 (2012).
- Kim, J. M., Rasmussen, J. P. & Rudensky, A. Y. Regulatory T cells prevent catastrophic autoimmunity throughout the lifespan of mice. *Nature Immunol.* **8**, 191–197 (2007).
- Erlebacher, A., Vencato, D., Price, K. A., Zhang, D. & Glimcher, L. H. Constraints in antigen presentation severely restrict T cell recognition of the allogeneic fetus. *J. Clin. Invest.* **117**, 1399–1411 (2007).
- Ertelt, J. M. *et al.* Selective priming and expansion of antigen-specific Foxp3[−] CD4⁺ T cells during *Listeria monocytogenes* infection. *J. Immunol.* **182**, 3032–3038 (2009).
- Pepper, M. *et al.* Different routes of bacterial infection induce long-lived T_H1 memory cells and short-lived T_H17 cells. *Nature Immunol.* **11**, 83–89 (2010).
- Koch, M. A. *et al.* The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation. *Nature Immunol.* **10**, 595–602 (2009).
- Wakim, L. M. & Bevan, M. J. From the thymus to longevity in the periphery. *Curr. Opin. Immunol.* **22**, 274–278 (2010).
- Trupin, L. S., Simon, L. P. & Eskenazi, B. Change in paternity: a risk factor for preeclampsia in multiparas. *Epidemiology* **7**, 240–244 (1996).
- Conde-Agudelo, A. & Belizan, J. M. Maternal morbidity and mortality associated with interpregnancy interval: cross sectional study. *BMJ* **321**, 1255–1259 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Jenkins for providing 2W1S-expressing mice, and A. Rudensky for providing *Foxp3*^{DTR} mice. This research was supported by NIH-NIAID awards R01AI087830 and R01AI100934 (S.S.W.), and NIH-NIDDK award F30DK084674 (J.H.R.). S.S.W. holds an Investigator in the Pathogenesis of Infectious Disease award from the Burroughs Wellcome Fund.

Author Contributions J.H.R., J.M.E., L.X. and S.S.W. designed and performed the experiments, and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.S.W. (singsing.way@cchmc.org).

METHODS

Mice. C57BL/6 (B6; H-2^{b/b}) or CD45.1 (H-2^{b/b}) and CD90.1 (H-2^{b/b}) mice on the B6 background, and Balb/c (H-2^{d/d}) mice, were purchased from The Jackson Laboratory or The National Cancer Institute. Transgenic mice expressing 2W1S_{55–68} antigen behind the β -actin promoter in all cells, and *Foxp3*^{DTR/DTR} mice where T_{reg} cells become susceptible to ablation with low-dose diphtheria toxin have each been described^{16,22}. *Foxp3*^{DTR/DTR} mice on the B6 background were intercrossed with CD45.1^{+/+} *Foxp3*^{DTR/DTR} mice. For mating, 2W1S-expressing males were used either on the B6 background, or backcrossed five generations to Balb/c mice, and verified to be H-2^{d/d}. Males were introduced to either virgin or post-partum (>14 days) females for 24 h, and mated mice visualized by a copulation plug representing E0.5. In each experiment, pups were removed within 24 h after delivery to prevent the potential transfer of fetal antigen through suckling. For sterilization, female mice were sub-lethally irradiated (100 rads) before mating.

Tetramer enrichment and flow cytometry. Phycoerythrin or allophycocyanin-conjugated MHC class II I-A^b 2W1S_{55–68} tetramer, and their use with anti-fluorophore-conjugated magnetic beads (Miltenyi Biotec) for enrichment have been described^{6,30}. For enumerating total 2W1S⁺ cells, all nucleated cells from secondary lymphoid tissue (spleen and axillary, brachial, cervical, inguinal, mesenteric, pancreatic, para-aortic/uterine lymph nodes) were collected, enriched using I-A^b 2W1S_{55–68} tetramer, and stained for cell-surface CD4, CD8 α , CD25, CD45.1, CD45.2, CD90.1, CD90.2, CD44, CD11b, CD11c, B220, F4/80, intracellular IFN- γ , IL-4, IL-10, IL-17A, or intranuclear FOXP3, Helios, Ki67, or T-bet expression using commercially available antibodies and cell permeabilization reagents (BD Pharmingen or eBioscience). For stimulation, cells were cultured with phorbol myristate acetate/ionomycin for 5 h in media supplemented with brefeldin A before tetramer staining.

Cell transfer and ablation. For adoptive transfer, CD4 cells in the spleen and lymph nodes were first purified by negative selection (Miltenyi Biotec), and one mouse equivalent of donor cells intravenously transferred into recipient mice 1 day before mating and/or infection. For ablation of donor *Foxp3*^{DTR/DTR} cells, recipient *Foxp3*^{WT/WT} mice were treated with purified diphtheria toxin (Sigma Chemicals, two doses 8 h apart, 0.5 μ g per dose). For ablation of endogenous T_{reg} cells in *Foxp3*^{DTR/WT} or *Foxp3*^{DTR/DTR} mice during pregnancy, purified diphtheria toxin was administered daily (0.5 μ g first dose, followed by 0.1 μ g per dose thereafter beginning mid-gestation) for 4 consecutive days. For FOXP3⁺ cell reconstitution requiring sustained ablation of endogenous T_{reg} cells in *Foxp3*^{DTR/DTR} recipient mice, purified diphtheria toxin was administered daily (0.5 μ g first dose, followed by 0.1 μ g per dose thereafter) immediately following transfer of purified donor CD4 cells as described².

Bacteria. *Listeria monocytogenes* was engineered to stably express and secrete 2W1S_{55–68} antigen by sub-cloning the *hly* promoter, signal sequence, 2W1S_{55–68} and ovalbumin coding sequence from the pAM401-based expression construct^{24,31} into the temperature sensitive plasmid pKSV7³². This pKSV7-based plasmid containing the *hly* promoter, signal sequence, 2W1S_{55–68} and ovalbumin coding sequence was used to electroporate Lm-OVA³³, and transformants selected by resistance to chloramphenicol (10 μ g ml⁻¹ final concentration) at room temperature. Individual clones were then passaged five times at 40 °C in brain heart infusion media (Becton Dickinson) with chloramphenicol selection (plasmid integration into the *L. monocytogenes* genome), followed by passage without antibiotic selection (plasmid excision from the *L. monocytogenes* genome). *L. monocytogenes* clones where double-homologous recombination had occurred were initially screened by replica plating for sensitivity to chloramphenicol, and a single clone (Lm-2W1S) verified by PCR and DNA sequencing using previously described methods³⁴. For infection, Lm-2W1S was grown to early log phase (OD_{600 nm} 0.1) in brain heart infusion media, washed and diluted with sterile saline, and then inoculated intravenously in the lateral tail vein (10⁴ colony-forming units per mouse). Five days thereafter, the antigen-specific CD4 cell response in the spleen and lymph nodes was investigated with I-A^b 2W1S_{55–68} tetramer staining.

Data acquisition and analysis. Cells stained with fluorochrome-conjugated tetramer and antibody were acquired on a FACSCanto cytometer (Becton Dickinson), and analysed using FlowJo (TreeStar) software. The number and percent cells were then analysed and found to be normally distributed, and thereafter the difference between separate groups of mice were analysed using an unpaired Student's *t* test, whereas differences between individual cell subsets within the same mouse were analysed using the paired Student's *t* test (Prism, GraphPad). For all analysis, *P* < 0.05 was taken as statistically significant.

30. Moon, J. J. *et al.* Tracking epitope-specific T cells. *Nature Protocols* **4**, 565–581 (2009).
31. Wirth, R., An, F. Y. & Clewell, D. B. Highly efficient protoplast transformation system for *Streptococcus faecalis* and a new *Escherichia coli*-*S. faecalis* shuttle vector. *J. Bacteriol.* **165**, 831–836 (1986).
32. Smith, K. & Youngman, P. Use of a new integrational vector to investigate compartment-specific expression of the *Bacillus subtilis* *spoIIIM* gene. *Biochimie* **74**, 705–711 (1992).
33. Foulds, K. E. *et al.* Cutting edge: CD4 and CD8 T cells are intrinsically different in their proliferative responses. *J. Immunol.* **168**, 1528–1532 (2002).
34. Rudd, B. D. *et al.* Nonrandom attrition of the naive CD8⁺ T-cell pool with aging governed by T-cell receptor:pMHC interactions. *Proc. Natl Acad. Sci. USA* **108**, 13694–13699 (2011).

Rapid induction of inflammatory lipid mediators by the inflammasome *in vivo*

Jakob von Moltke¹, Norver J. Trinidad¹, Mahtab Moayeri⁴, Alexander F. Kintzer², Samantha B. Wang³, Nico van Rooijen⁵, Charles R. Brown⁶, Bryan A. Krantz², Stephen H. Leppla⁴, Karsten Gronert³ & Russell E. Vance¹

Detection of microbial products by host inflammasomes is an important mechanism of innate immune surveillance. Inflammasomes activate the caspase-1 (CASP1) protease, which processes the cytokines interleukin (IL)-1 β and IL-18, and initiates a lytic host cell death called pyroptosis¹. To identify novel CASP1 functions *in vivo*, we devised a strategy for cytosolic delivery of bacterial flagellin, a specific ligand for the NAIP5 (NLR family, apoptosis inhibitory protein 5)/NLRC4 (NLR family, CARD-domain-containing 4) inflammasome^{2–4}. Here we show that systemic inflammasome activation by flagellin leads to a loss of vascular fluid into the intestine and peritoneal cavity, resulting in rapid (less than 30 min) death in mice. This unexpected response depends on the inflammasome components NAIP5, NLRC4 and CASP1, but is independent of the production of IL-1 β or IL-18. Instead, inflammasome activation results, within minutes, in an ‘eicosanoid storm’—a pathological release of signalling lipids, including prostaglandins and leukotrienes, that rapidly initiate inflammation and vascular fluid loss. Mice deficient in cyclooxygenase-1, a critical enzyme in prostaglandin biosynthesis, are resistant to these rapid pathological effects of systemic inflammasome activation by either flagellin or anthrax lethal toxin. Inflammasome-dependent biosynthesis of eicosanoids is mediated by the activation of cytosolic phospholipase A₂ in resident peritoneal macrophages, which are specifically primed for the production of eicosanoids by high expression of eicosanoid biosynthetic enzymes. Our results therefore identify eicosanoids as a previously unrecognized cell-type-specific signalling output of the inflammasome with marked physiological consequences *in vivo*.

NAIP/NLRC4 inflammasome activation is critical for innate immune detection and defence against multiple bacterial pathogens in mice^{5–7}. This resistance to infection, as well as the inflammasome-dependent response to systemic endotoxin, does not require IL-1 β or IL-18 (refs 5, 8, 9), suggesting a critical role for pyroptosis and/or other inflammasome functions^{9–12}. In this study we sought to identify previously unknown *in vivo* signalling outputs of the NAIP5/NLRC4 inflammasome. To activate NAIP5/NLRC4 selectively, we delivered *Legionella pneumophila* flagellin (FlaA) to the cytosol by the fusion of FlaA to the amino-terminal domain of *Bacillus anthracis* lethal factor (LFn), which mediates cytosolic delivery through the anthrax protective antigen (PA) channel¹³. As expected, PA plus LFn–FlaA (here called FlaTox), but not PA or LFn–FlaA alone, activated the NAIP5/NLRC4 inflammasome in bone-marrow-derived macrophages (BMDMs), as indicated by cleavage of CASP1, release of lactate dehydrogenase and secretion of IL-1 β (Supplementary Fig. 1). Inflammasome activation was abrogated in *Casp1*^{–/–}, *Naip5*^{–/–} and *Nlrc4*^{–/–} BMDMs (Supplementary Fig. 1b). A mutant form of FlaTox (FlaTox(AAA)), which is recognized by Toll-like receptor (TLR)-5 but not by NAIP5 (ref. 4), did not activate pyroptosis (Supplementary Fig. 1a).

We then determined the effect of FlaTox administration *in vivo*. Intravenous or intraperitoneal delivery of FlaTox killed mice rapidly,

causing symptoms within 15 min and a mean time to death of about 30 min at saturating intravenous doses (Fig. 1a and Supplementary Fig. 1f). For subsequent experiments we administered a sublethal intraperitoneal dose. FlaTox-treated mice rapidly developed diarrhoea; however, histological analysis (Supplementary Fig. 2 and data not shown) after 30 min revealed no signs of pathology. Instead, we found fluid accumulation in the peritoneal cavity and intestine, but not in the kidneys or lungs (Supplementary Fig. 3). This fluid was lost from the blood, as the percentage volume of red blood cells (haematocrit) increased to 70–80% (normal 45–50%) within 30–40 min after FlaTox injection (Fig. 1b). As haematocrit rose, body temperature dropped, but with delayed kinetics (Fig. 1b). Haemoconcentration was the earliest pathological event that we detected after treatment with FlaTox, and is probably the primary cause of the ensuing circulatory collapse, hypothermia and death.

We found that only the complete toxin, rather than individual subunits or FlaTox(AAA), induced hypothermia and increased haematocrit (Fig. 1c, d and Supplementary Fig. 4a), showing that TLR signalling activated by flagellin or bacterial contaminants is insufficient to cause pathology. *Nlrc4*^{–/–} mice were completely protected at all doses and time points tested (Fig. 1e, f and data not shown). *Naip5*^{–/–} and *Casp1*^{–/–} mice were completely protected in survival experiments (Supplementary Fig. 4b), but they showed a delayed haemoconcentration and hypothermic response (Fig. 1e, f). The moderate response of *Naip5*^{–/–} mice may be due to recognition of flagellin by NAIP6 (refs 2, 3). Mice lacking tumour necrosis factor (*LT α /LT β /TNF α* ^{–/–}) or IL-1 β and IL-18 were as sensitive as wild-type (B6) mice (Fig. 1g, h and Supplementary Fig. 4b–d), ruling out an essential role for these cytokines in FlaTox-induced pathologies.

To identify the cell type(s) that respond to FlaTox, we generated bone marrow chimaeras by using susceptible wild-type (B6) mice and resistant *Nlrc4*^{–/–} mice. *Nlrc4*^{–/–} mice reconstituted with wild-type bone marrow were completely susceptible to FlaTox (Fig. 2a, b). Wild-type mice reconstituted with *Nlrc4*^{–/–} bone marrow also responded to FlaTox, but with delayed kinetics. These results suggest that at least two cell populations respond to FlaTox: first, radio-sensitive haematopoietic cells that are necessary and sufficient for the early response to FlaTox (0–30 min), and second, radio-resistant cells that respond after 30 min. Hereafter we focused on the early haematopoietic response (EHR).

The EHR was intact in mice deficient in mast cells, lymphocytes and neutrophils (Supplementary Fig. 5). By contrast, wild-type mice depleted of macrophages by using clodronate-loaded liposomes were almost completely protected from the EHR (Fig. 2c, d). Similarly, depletion of CD11b⁺ cells in FVB:CD11b-DTR mice conferred protection (Fig. 2e, f). Flow cytometric analysis revealed almost complete ablation of CD11b/F4-80^{hi} resident peritoneal macrophages by both treatments (Supplementary Fig. 6a), whereas depletion of splenic and lamina propria macrophages was only partial (clodronate) or not

¹Department of Molecular and Cell Biology, Division of Immunology and Pathogenesis, University of California at Berkeley, Berkeley, California 94720, USA. ²Department of Chemistry, University of California at Berkeley, Berkeley, California 94720, USA. ³Vision Science Program, School of Optometry, University of California at Berkeley, Berkeley, California 94720, USA. ⁴Laboratory of Parasitic Diseases, Microbial Pathogenesis Section, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁵Department of Molecular Cell Biology, Faculty of Medicine, Vrije Universiteit, Amsterdam 1081 BT, The Netherlands. ⁶Department of Veterinary Pathobiology, University of Missouri, Columbia, Missouri 65211, USA.

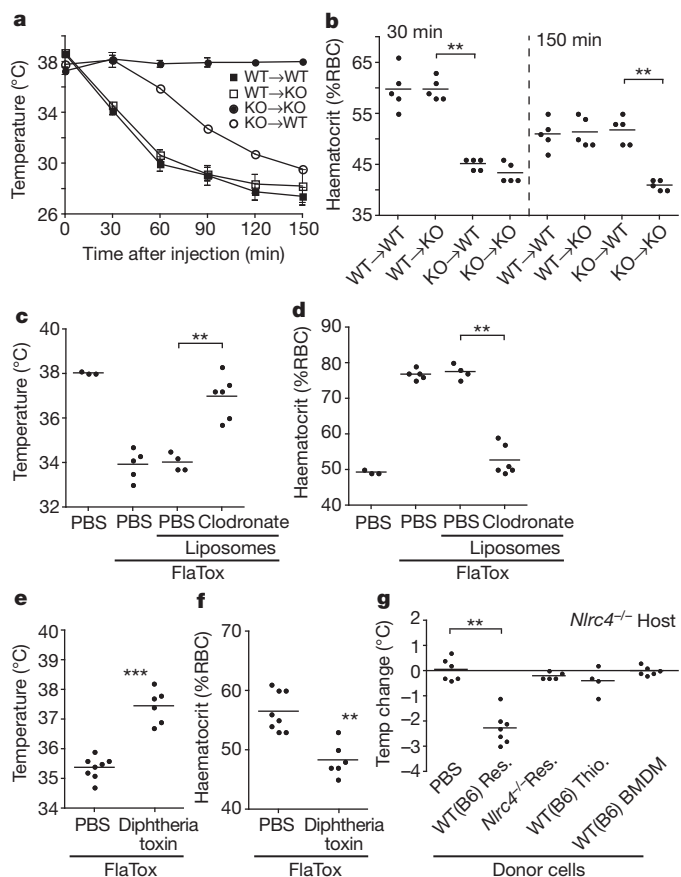


Figure 2 | Resident peritoneal macrophages are critical for the early FlaTox response *in vivo*. a–f, Mice were injected intraperitoneally with FlaTox, and rectal temperature (a, c, e) or haematocrit (b, d, f) were measured after 30 min or at the indicated times. a, b, Bone marrow chimeric mice (KO = *Nlr4*^{-/-}; *n* = 5). c, d, Macrophage-depleted wild-type (B6) mice (*n* = 3–6). e, f, CD11b⁺ cell-depleted FVB:CD11b-DTR mice (*n* = 6–7). g, *Nlr4*^{-/-} host mice injected intraperitoneally with 10⁷ resident (Res.) or thioglycollate-elicited (Thio.) peritoneal cells or BMDMs of indicated genotype. Rectal temperature was measured 30 min after intraperitoneal injection of FlaTox (8 μg g⁻¹ PA + 4 μg g⁻¹ LFn-FlaA). Data shown are pooled from multiple experiments (g) or are representative of at least three independent experiments (a–f). Error bars in a denote s.e.m. Two asterisks, *P* < 0.01; three asterisks, *P* = 0.001 (Mann-Whitney test).

chemotaxis^{13–16}. Eicosanoids are synthesized when arachidonic acid released from cell membranes by phospholipases is converted into prostaglandins (PGs) and thromboxanes downstream of the cyclooxygenases (COX-1 and COX-2) or into hydroxyeicosatetraenoic acids (HETEs) and leukotrienes (LTs) downstream of the lipoxygenases (12/15-LOX and 5-LOX) (Supplementary Fig. 7). Direct intraperitoneal injection of prostaglandins leads to diarrhoea and fluid accumulation in the gut^{17,18}, both hallmarks of FlaTox-induced pathology.

We therefore speculated that resident peritoneal macrophages, the specific cells required for the EHR (Fig. 2), produce eicosanoids in response to inflammasome activation. Using liquid chromatography tandem mass spectrometry (LC-MS/MS) lipidomic analysis and enzyme immunoassay, we detected rapid biosynthesis of numerous COX-dependent and LOX-dependent eicosanoids including LTB₄, PGE₂ and cysteinyl LTs in supernatants of peritoneal lavage cells treated *ex vivo* with FlaTox, whereas the FlaTox(AAA) mutant elicited a much weaker response (Fig. 3 and Supplementary Fig. 8a, b). Moreover, eicosanoid induction required NAIP5, NLRC4 and CASP1 (Fig. 3b and Supplementary Fig. 8c). LTB₄/PGE₂ induction by PA or LFn-FlaA alone was equivalent to stimulation with FlaTox(AAA) or lipopolysaccharide, further demonstrating that TLR signalling cannot account for their rapid biosynthesis (Fig. 3c and Supplementary

Fig. 8b). The flagellated intracellular pathogen *Salmonella enterica* serovar Typhimurium (*S. typhimurium*) also elicited inflammasome-dependent eicosanoid biosynthesis in a similar manner to FlaTox (Fig. 3d and Supplementary Fig. 8d), demonstrating that this pathway is activated during live infection.

Similarly to the *ex vivo* results, LC-MS/MS analysis of peritoneal lavage fluid from mice treated with FlaTox for 20 min revealed robust inflammasome-dependent eicosanoid biosynthesis *in vivo* (Supplementary Fig. 9). Some residual eicosanoid biosynthesis (particularly of PGE₂) was observed after treatment with FlaTox(AAA). This residual response is probably due to bacterial contaminants (for example lipopolysaccharide) activating peritoneal cells other than macrophages, because macrophages did not produce PGE₂ in response to stimulation for 30 min with lipopolysaccharide or FlaTox(AAA) (Fig. 3c). However, because this residual response did not produce symptoms (Fig. 1c, d) and eicosanoids have only paracrine effects, the FlaTox-induced pathology is probably a localized response mediated by resident peritoneal macrophages and/or the synergistic effects of multiple eicosanoids. FlaTox did not result in the detectable production of anti-inflammatory lipoxins¹⁹. Taken together, our results show that inflammasome activation results in an 'eicosanoid storm' *ex vivo* and *in vivo*, characterized by the broad biosynthesis of both LOX and COX products, with a bias towards pro-inflammatory lipids.

How does inflammasome activation induce eicosanoid biosynthesis? Although cells express several initiating A₂ phospholipases,

Ca²⁺-dependent cytosolic phospholipase A₂ (cPLA₂) accounts for nearly all eicosanoid biosynthesis in peritoneal macrophages treated with a Ca²⁺ ionophore²⁰. Inhibition of cPLA₂ also blocked LTB₄ and PGE₂ production but not pyroptosis in response to FlaTox (Fig. 4a and Supplementary Fig. 8e). An increase in intracellular Ca²⁺ is both necessary and sufficient for cPLA₂ activation²¹; notably, the earliest detectable CASP1-dependent events in BMDMs infected with *S. typhimurium* are the formation of plasma membrane pores and the influx of Ca²⁺ (refs 22, 23). We observed that CASP1-dependent Ca²⁺ influx, comparable in magnitude to an ionomycin control (Supplementary Fig. 10), preceded cell lysis (79 ± 29 s before the onset of membrane blebbing) in resident peritoneal macrophages (Fig. 4b, c). This Ca²⁺ influx seems to be critical for eicosanoid biosynthesis but dispensable for pyroptosis in response to FlaTox, because the intracellular Ca²⁺ chelator bis-(*o*-aminophenoxy)ethane-*N,N,N',N'*-tetraacetic acid acetoxymethyl ester (BAPTA-AM) inhibited PGE₂ and LTB₄ production without blocking lactate dehydrogenase (LDH) release (Fig. 4d and Supplementary Fig. 8f). Macrophages treated with glycine to inhibit pyroptosis²³ still produced PGE₂ in response to FlaTox, whereas cells lysed by digitonin or H₂O₂ did not, further indicating that eicosanoid production and cell lysis are separable events (Supplementary Fig. 11). Taken together, our results suggest a model (Supplementary Fig. 12) in which CASP1 activation results in rapid Ca²⁺ flux (Fig. 4b) that leads to cPLA₂ activation and downstream eicosanoid biosynthesis. Identification of a CASP1 substrate required for Ca²⁺ influx is an important, but technically challenging, area for future research because proteomic studies to identify novel substrates of CASP1 (refs 24–26) have so far yielded limited insight.

Although cPLA₂ activity is regulated primarily by Ca²⁺ influx, its activity can be enhanced by mitogen-activated protein kinase (MAPK)-dependent phosphorylation²¹. We wondered whether TLR signalling activated by FlaTox (FlaA or bacterial contaminants) might enhance cPLA₂ activity through downstream MAPKs. Indeed, cPLA₂ in resident peritoneal cells was phosphorylated after treatment with FlaTox, and this was inflammasome-independent but *Myd88/Trif*-dependent (Fig. 4e). Accordingly, LTB₄ production in response to FlaTox was partly attenuated in *Myd88/Trif*^{-/-} cells, and *Myd88/Trif*^{-/-} mice were partly protected *in vivo* (Supplementary Fig. 13). Thus, although inflammasome-dependent Ca²⁺ flux is both sufficient and necessary for eicosanoid production in response to FlaTox, TLR signalling, which is expected to accompany natural infection, can synergize with inflammasome activation to produce maximal responses.

We further investigated the basis for the cell-type specificity of inflammasome-dependent eicosanoid biosynthesis. Consistent with the inability of BMDMs to transfer responsiveness to FlaTox *in vivo* (Fig. 2g), we observed no eicosanoid production by LC-MS/MS in BMDMs treated with FlaTox for 30 min (Fig. 3a). Even at 2 h after treatment, when most BMDMs have undergone pyroptosis, we could not detect PGE₂ or LTB₄ production (Fig. 4f, g and Supplementary Fig. 8g, h). *Cox1*, *Alox12/15* and *Alox5*, which encode key enzymes required for eicosanoid biosynthesis, are expressed at much higher levels (10–1,000-fold) in CD11b/F4-80^{hi} resident peritoneal macrophages than in BMDMs or thioglycollate-elicited CD11b/F4-80^{hi} cells (Fig. 4h–j). Resident peritoneal macrophages are therefore uniquely primed for eicosanoid responses. Most characterization of inflammasomes has relied on BMDMs, which perhaps explains why a link to eicosanoid biosynthesis has not been described. We speculate that the primed state of resting peritoneal macrophages may be a general characteristic of resident macrophages guarding sites of pathogen entry; it will be important to further explore the lineage-specific regulation of inflammasome function *in vivo*.

To test whether eicosanoids contribute to FlaTox-induced pathology, we injected B6;129P2.*Cox1*^{-/-} mice and littermate controls with FlaTox, and found that the EHR was significantly attenuated in these mice (*P* < 0.0007; Fig. 4k, l) despite intact pyroptosis (Supplementary Fig. 14). As expected, given that FlaTox induces the broad biosynthesis

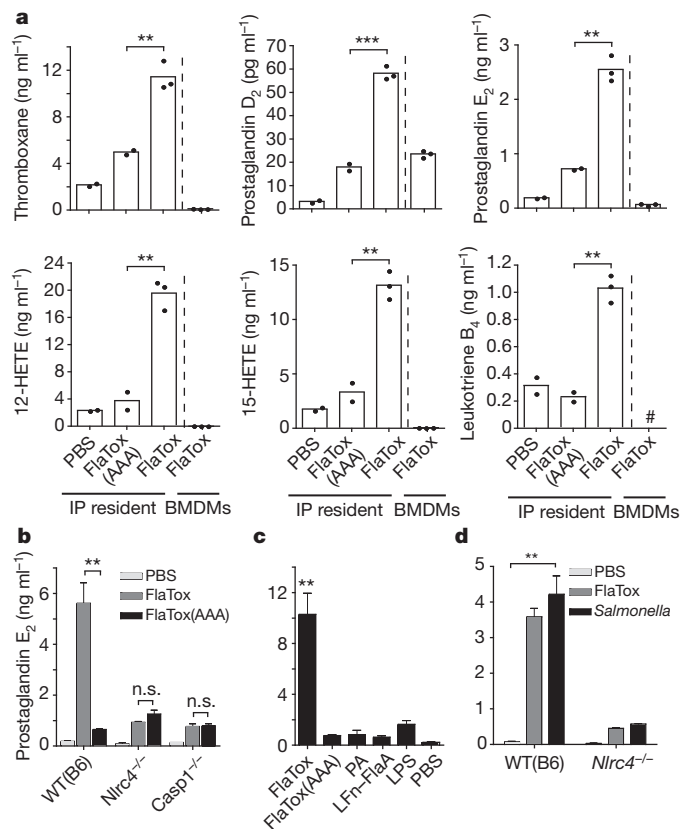


Figure 3 | Inflammasome-dependent eicosanoid biosynthesis. **a**, LC-MS/MS-based lipidomics of wild-type (B6) BMDMs or resident peritoneal (IP) cells incubated for 30 min *ex vivo* with FlaTox (20 µg ml⁻¹ PA + 10 µg ml⁻¹ LFN-FlaA). **b–d**, PGE₂ immunoassay of resident peritoneal macrophages treated *ex vivo* with FlaTox, the indicated proteins (10 µg ml⁻¹ PA in **d**; 5 µg ml⁻¹ in all others), lipopolysaccharide (**c**, 1 µg ml⁻¹) or *S. typhimurium* (**d**, multiplicity of infection = 5) and incubated for 30 min (**b**, **c**) or 180 min (**d**). Data shown are representative of at least two (**d**) or three (**a–c**) independent experiments. Error bars in **b–d** denote s.e.m. Two asterisks, *P* < 0.01; three asterisks, *P* < 0.0005 (Student's *t*-test). Hash sign, not detected.

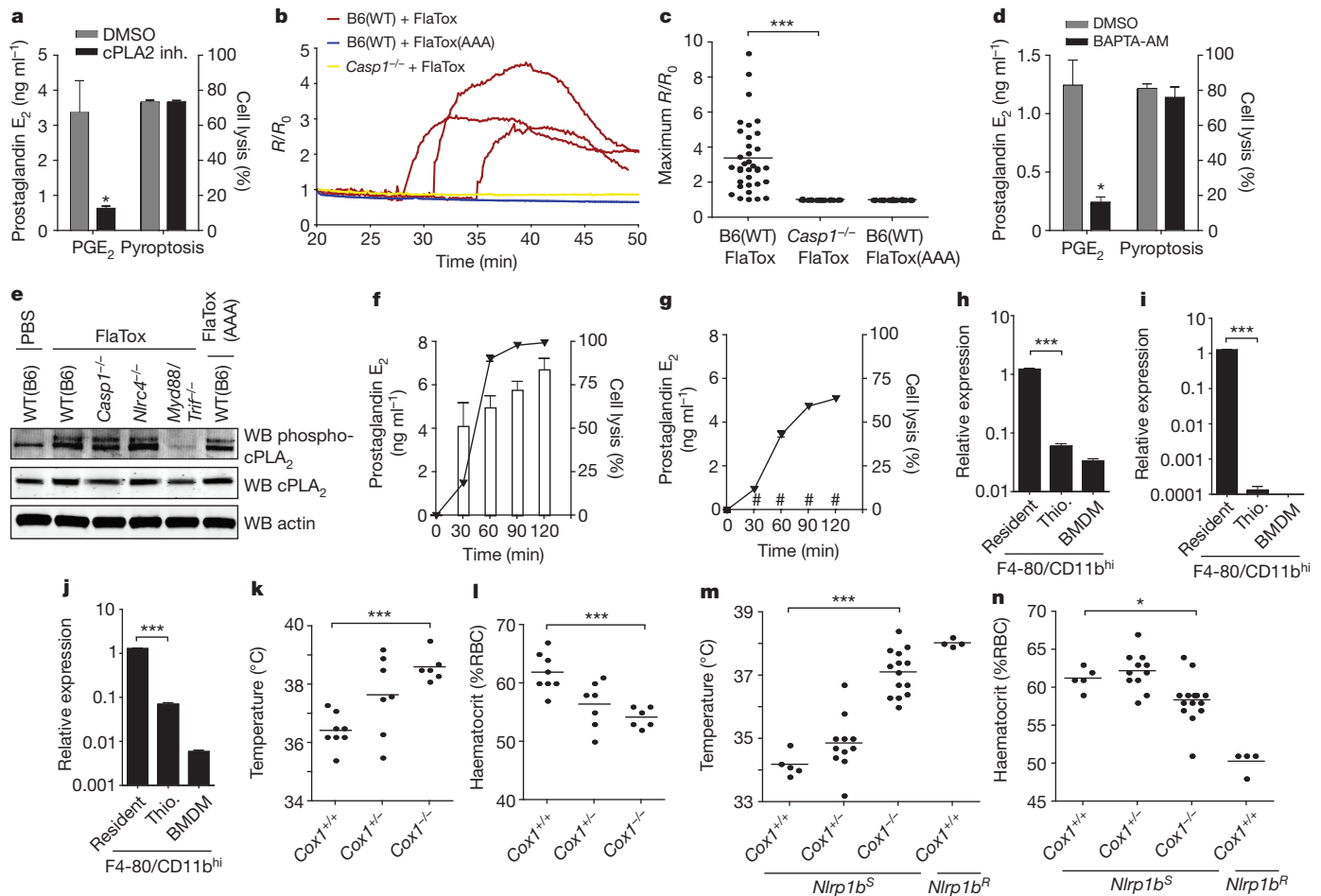


Figure 4 | Mechanism and *in vivo* role of eicosanoid production. **a, d,** PGE₂ immunoassay (30 min) or LDH assay (2 h) on supernatants from wild-type (B6) resident peritoneal macrophages pretreated for 45 min with dimethylsulphoxide (DMSO), cPLA₂ inhibitor (a, 0.2 μM pyrrophenone) or BAPTA-AM (d, 10 μM) then FlaTox + DMSO/inhibitor. **b, c,** Wild-type (B6) or *Casp1*^{-/-} resident peritoneal macrophages were treated with FlaTox or FlaTox(AAA) and Ca²⁺ indicator (Fluo-4; 2.5 mM) and fluorescence/background (*R/R*₀) was quantified over time. **b,** Representative cell traces. **c,** Maximum *R/R*₀ for each cell. **e,** Resident peritoneal cells of indicated genotype treated for 30 min as indicated. Cell lysates were probed for indicated proteins by western blot (WB). **f, g,** Resident peritoneal macrophages (**f**) or BMDMs (**g**) treated with FlaTox. PGE₂ (bars) and cell lysis (line) were measured over time as in **a**. **h–j,** Expression of *Cox1* (**h**), *Alox12/15* (**i**) and *Alox5*

(**j**) measured by quantitative RT-PCR in BMDMs and resident peritoneal macrophages or thioglycollate-elicited peritoneal cells (thio.) sorted for CD11b/F4-80^{hi}. **k–n,** Mice were injected intraperitoneally with FlaTox (**k, l**) or intravenously with anthrax lethal toxin (**m, n**; 200 μg PA + 100 μg lethal factor); rectal temperature (**k, m**) and haematocrit (**l, n**) were measured after 30 min (**k, l**) or 45 min (**m, n**). **k, l,** B6;129P2-*Cox1*^{-/-} mice and littermate controls. **m, n,** *Cox1*^{-/-} mice and littermate controls expressing a lethal-toxin-sensitive (S) or resistant (R) *Nlrp1b* allele. Data are pooled from multiple experiments (**k–n**) or are representative of at least two (**b, c, h–j**) or three (**a, d–g**) independent experiments. Error bars in **a, d** and **f–j** denote s.e. Asterisk, *P* < 0.04; two asterisks, *P* < 0.006; three asterisks, *P* < 0.0007 (**k–n**, Mann–Whitney test; **a, c, d, h–j**, Student's *t*-test). Hash sign, not detected.

of both COX-1-dependent and COX-1-independent eicosanoids, the role of COX-1 was masked at high doses of FlaTox or at later time points (data not shown), suggesting a contribution of COX-1-independent eicosanoids or other unidentified factors. In confirmation of the results with *Cox1*^{-/-} mice, chemical inhibition of COX-1 protected mice from a low dose of FlaTox (Supplementary Fig. 15). Inhibition of the inducible enzyme COX-2 or the genetic deletion of *Alox5* had no effect, although their contribution may be masked by functional redundancy of downstream eicosanoids (Supplementary Fig. 15). These data link the production of COX-1-dependent eicosanoids to the pathology associated with *in vivo* delivery of FlaTox, which is consistent with earlier reports that purified prostaglandins are sufficient to cause fluid accumulation and diarrhoea^{17,18}.

We speculated that other inflammasomes might also lead to eicosanoid production *in vivo*. For example, anthrax lethal toxin activates the NLRP1B inflammasome²⁷, resulting in a rapid (but transient and non-lethal) hypothermic response²⁸. Because C57BL/6 mice express a non-functional allele of *Nlrp1b* (*Nlrp1b*^R), we injected *Cox1*^{+/+} mice expressing the sensitive 129 allele of *Nlrp1b* (*Nlrp1b*^S).

These mice developed diarrhoea, decreased body temperature and an inflammasome-dependent increase in haematocrit (Fig. 4m, n). As in FlaTox-treated mice, deletion of *Cox1* attenuated these early pathologies of lethal toxin (Fig. 4m, n), indicating that this naturally occurring bacterial toxin activates a similar, though non-lethal, inflammasome pathway in mice.

Whereas many cellular immune responses require *de novo* transcription, the NAIP5/NLRC4 and NLRP1B inflammasomes assemble from preformed protein components to activate a proteolytic cascade. These inflammasomes are therefore ideally positioned to mediate rapid responses to infection. Initiated within minutes of flagellin detection, the inflammasome-dependent eicosanoid production described here is one of the most rapid innate immune cellular responses known *in vivo*. Although performed *in vivo*, the results presented here rely on the systemic administration of toxins. It will be important to explore the role of inflammasome-dependent eicosanoids during live infection. When restricted to the site of infection, such eicosanoids may have a beneficial role in host defence, for example by rapidly increasing local vascular permeability, allowing an influx of antibody,

complement and immune cells. Future studies should also evaluate a role for eicosanoids in other inflammasome-dependent phenotypes previously described *in vivo*. Indeed, our results suggest that the signalling outputs of inflammasomes may be much broader than has previously been realized.

METHODS SUMMARY

Toxin delivery and pathology. Recombinant proteins (PA, LFn-FlaA, LFn-FlaA(AAA) and LF) were purified from *Escherichia coli* as described previously²⁹, and endotoxin was removed with Detoxi-Gel (Pierce). LFn-FlaA(AAA) was generated by mutating the three carboxy-terminal leucines of *L. pneumophila* flagellin to alanine⁴. Unless otherwise noted, standard toxin doses were 2 µg g⁻¹ body weight of LFn-FlaA *in vivo* (200 µl intraperitoneally) and 5 µg ml⁻¹ LFn-FlaA *in vitro*. PA dose was always 2 × LFn-FlaA. Rectal temperature was measured with a MicroTherma 2T thermometer (Braintree Scientific). Blood for haematocrit was collected by retro-orbital bleed into StatSpin microhaematocrit tubes (Fisher Scientific).

Mice. For bone marrow chimaeras, mice were irradiated twice with 600 rad 4 h apart, injected with 5 × 10⁶ donor cells, and analysed after 12–15 weeks. Macrophages were depleted for 48 h with liposome-encapsulated clodronate (N.v.R.) (500 µl intraperitoneally and 200 µl intravenously). CD11b⁺ cells were depleted for 24 h in FVB-Tg(CD11b/EGFP)34Lan/J mice with 25 ng g⁻¹ body weight (intraperitoneally) of diphtheria toxin (Sigma-Aldrich).

Lipidomics. Lipid autacoids were extracted by solid phase with SampliQ ODS-C₁₈ cartridges (Agilent Technologies). Eicosanoids and docosanoids were identified and quantified with a triple-quadrupole linear ion-trap LC–MS/MS system (MDS SCIEX 3200 QTRAP) equipped with a Kinetex C₁₈ mini-bore column.

Statistical analysis. Statistical differences were calculated with an unpaired two-tailed Student's *t*-test (*in vitro/ex vivo*) or two-tailed non-parametric Mann–Whitney test (*in vivo*) using GraphPad Prism 4.0b.

Full Methods and any associated references are available in the online version of the paper.

Received 6 December 2011; accepted 26 June 2012.

Published online 19 August 2012.

- Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821–832 (2010).
- Kofoed, E. M. & Vance, R. E. Innate immune recognition of bacterial ligands by NALPs determines inflammasome specificity. *Nature* **477**, 592–595 (2011).
- Zhao, Y. *et al.* The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature* **477**, 596–600 (2011).
- Lightfield, K. L. *et al.* Critical function for Naip5 in inflammasome activation by a conserved carboxy-terminal domain of flagellin. *Nature Immunol.* **9**, 1171–1178 (2008).
- Case, C. L., Shin, S. & Roy, C. R. Asc and Ipaf inflammasomes direct distinct pathways for caspase-1 activation in response to *Legionella pneumophila*. *Infect. Immun.* **77**, 1981–1991 (2009).
- Sutterwala, F. S. *et al.* Immune recognition of *Pseudomonas aeruginosa* mediated by the IPAF/NLRC4 inflammasome. *J. Exp. Med.* **204**, 3235–3245 (2007).
- Miao, E. A. *et al.* Innate immune detection of the type III secretion apparatus through the NLRC4 inflammasome. *Proc. Natl Acad. Sci. USA* **107**, 3076–3080 (2010).
- Miao, E. A. *et al.* Caspase-1-induced pyroptosis is an innate immune effector mechanism against intracellular bacteria. *Nature Immunol.* **11**, 1136–1142 (2010).
- Lamkanfi, M. *et al.* Inflammasome-dependent release of the alarmin HMGB1 in endotoxemia. *J. Immunol.* **185**, 4385–4392 (2010).
- Gurcel, L., Abrami, L., Girardin, S., Tschopp, J. & van der Goot, F. G. Caspase-1 activation of lipid metabolic pathways in response to bacterial pore-forming toxins promotes cell survival. *Cell* **126**, 1135–1145 (2006).
- Amer, A. *et al.* Regulation of *Legionella* phagosome maturation and infection through flagellin and host Ipaf. *J. Biol. Chem.* **281**, 35217–35223 (2006).
- Keller, M., Ruegg, A., Werner, S. & Beer, H. D. Active caspase-1 is a regulator of unconventional protein secretion. *Cell* **132**, 818–831 (2008).
- Funk, C. D. Prostaglandins and leukotrienes: advances in eicosanoid biology. *Science* **294**, 1871–1875 (2001).
- Samuelsson, B. From studies of biochemical mechanism to novel biological mediators: prostaglandin endoperoxides, thromboxanes, and leukotrienes. (Nobel Lecture, 8 December 1982.) *Biosci. Rep.* **3**, 791–813 (1983).
- Tobin, D. M. *et al.* The *Ita4h* locus modulates susceptibility to mycobacterial infection in zebrafish and humans. *Cell* **140**, 717–730 (2010).
- Serhan, C. N. & Haeggström, J. Z. in *Fundamentals of Inflammation* (eds Serhan, C. N., Ward, P. A. & Gilroy, D. W.) 153–175 (Cambridge Univ. Press, 2011).
- Robert, A., Nezamis, J. E., Lancaster, C., Hanchar, A. J. & Klepper, M. S. Enteropooling assay: a test for diarrhea produced by prostaglandins. *Prostaglandins* **11**, 809–828 (1976).
- Riviere, P. J., Farmer, S. C., Burks, T. F. & Porreca, F. Prostaglandin E₂-induced diarrhea in mice: importance of colonic secretion. *J. Pharmacol. Exp. Ther.* **256**, 547–552 (1991).
- Serhan, C. N. *et al.* Resolution of inflammation: state of the art, definitions and terms. *FASEB J.* **21**, 325–332 (2007).
- Bonventre, J. V. *et al.* Reduced fertility and postischemic brain injury in mice deficient in cytosolic phospholipase A₂. *Nature* **390**, 622–625 (1997).
- Gijon, M. A., Spencer, D. M. & Leslie, C. C. Recent advances in the regulation of cytosolic phospholipase A₂. *Adv. Enzyme Regul.* **40**, 255–268 (2000).
- Bergsbaken, T., Fink, S. L., den Hartigh, A. B., Loomis, W. P. & Cookson, B. T. Coordinated host responses during pyroptosis: caspase-1-dependent lysosome exocytosis and inflammatory cytokine maturation. *J. Immunol.* **187**, 2748–2754 (2011).
- Fink, S. L. & Cookson, B. T. Caspase-1-dependent pore formation during pyroptosis leads to osmotic lysis of infected host macrophages. *Cell. Microbiol.* **8**, 1812–1825 (2006).
- Lamkanfi, M. *et al.* Targeted peptidocentric proteomics reveals caspase-7 as a substrate of the caspase-1 inflammasomes. *Mol. Cell. Proteomics* **7**, 2350–2363 (2008).
- Shao, W., Yeretssian, G., Doiron, K., Hussain, S. N. & Saleh, M. The caspase-1 digestome identifies the glycolysis pathway as a target during infection and septic shock. *J. Biol. Chem.* **282**, 36321–36329 (2007).
- Agard, N. J., Maltby, D. & Wells, J. A. Inflammatory stimuli regulate caspase substrate profiles. *Mol. Cell. Proteomics* **9**, 880–893 (2010).
- Boyd, E. D. & Dietrich, W. F. *Nalp1b* controls mouse macrophage susceptibility to anthrax lethal toxin. *Nature Genet.* **38**, 240–244 (2006).
- Terra, J. K. *et al.* Cutting edge: resistance to *Bacillus anthracis* infection mediated by a lethal toxin sensitive allele of *Nalp1b/Nlrp1b*. *J. Immunol.* **184**, 17–20 (2010).
- Krantz, B. A. *et al.* A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore. *Science* **309**, 777–781 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Mariathasan and V. Dixit for *Nlrp4*^{-/-} mice; I. Bergin and S. Griffey for pathology reports; D. Crown for help with survival experiments; L. Lopez for support in our animal facility; D. Bautista and R. Nichols for help with calcium imaging; and M. Fontana and members of the Barton and Vance laboratories for discussions. Work in R.E.V.'s laboratory is supported by Investigator Awards from the Burroughs Wellcome Fund and the Cancer Research Institute and by National Institutes of Health (NIH) grants AI075039, AI080749 and AI063302. K.G.'s laboratory is supported by NIH grants EY016136 and EY022208. J.v.M. is supported by a grant from the Cancer Research Coordinating Committee of the University of California.

Author Contributions J.v.M. and R.E.V. conceived the study. J.v.M., R.E.V. and K.G. designed the experiments and wrote the paper. J.v.M. performed the experiments with help from N.J.T., M.M. performed experiments shown in Fig. 1a and Supplementary Fig. 1f. J.v.M., N.J.T., M.M., S.B.W., K.G. and R.E.V. analysed the results. A.F.K., B.A.K., C.R.B., S.H.L. and N.v.R. provided mice and/or reagents.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.E.V. (rvance@berkeley.edu) or K.G. (kgrobert@berkeley.edu).

METHODS

Mice and cell culture. Except bone marrow chimaeras (see below), all mice were sex and age-matched at 5–8 weeks old. C57BL/6J (B6) and B6.cKit^{wsh/wsh} mice were purchased from Jackson Laboratories. B6.129P2-Ptgs1^{tm1Unc} (Cox1) mice were purchased from Taconic. B6.Nlrp4^{-/-} mice³⁰ were from S. Mariathasan and V. Dixit. B6.Caspl1^{-/-} mice³¹ were a gift from A. Van der Velden and M. Starnbach. B6.Myd88/Trif^{-/-} and FVB-Tg(CD11b/EGFP)34Lan/J were a gift from G. Barton. B6.Rag1^{-/-} mice were a gift from D. Raulet. B6.II1b/II18^{-/-} mice were a gift from D. Portnoy. B6.Alox5^{-/-} mice were provided by C. Brown. B6.Naip5^{-/-} mice were described previously⁴. For bone marrow chimaeras, mice were irradiated twice with 600 rad 4 h apart and reconstituted with 5×10^6 donor cells by injection into the tail vein. Chimaeric mice were assayed 12–15 weeks after irradiation. Bone marrow macrophages were differentiated from bone-marrow-derived precursor cells by using macrophage colony stimulating factor as described previously⁴. For thioglycollate-elicited macrophages, mice were injected intraperitoneally with 2 ml of 4% aged thioglycollate medium (BD Diagnostics) 4 days before peritoneal lavage. Cell lysis was measured by LDH release assay (Promega) in accordance with the manufacturer's protocol. Animal experiments were approved by the Animal Care and Use Committee of the National Institute of Allergy and Infectious Diseases, National Institutes of Health (Fig. 1a and Supplementary Fig. 1f) and the University of California, Berkeley Animal Care and Use Committee (all other figures).

FlaTox injections and pathology. Recombinant proteins (PA, LFn-FlaA, LFn-FlaA(AAA) and LF) were purified from *E. coli* as described previously²⁹. Endotoxin was removed from these proteins with Detoxi-Gel (Pierce) in accordance with the manufacturer's protocol. LFn-FlaA(AAA) was generated by mutating the three carboxy-terminal leucines of *L. pneumophila* flagellin to alanine⁴. Toxin was injected intraperitoneally or intravenously (tail vein) in 200 μ l of PBS. Unless otherwise noted, standard doses were 2 μ g g⁻¹ body weight of LFn-FlaA *in vivo* (intraperitoneally) and 5 μ g ml⁻¹ LFn-FlaA *in vitro*. PA dose was always 2 \times LFn-FlaA. Rectal temperature was measured with a MicroTherma 2T thermometer (Braintree Scientific) with a lubricated RET-3 probe. For haematocrit measurement, mice were anaesthetized briefly with isoflurane and blood was collected in a heparinized StatSpin 40-mm tube (Fisher Scientific) by retro-orbital bleed. The tube was sealed at one end with StatSpin sealant (Fisher Scientific) and centrifuged for 10 min at 9,000g. The percentage of red blood cells was quantified with a StatSpin card haematocrit reader.

Fluid loss. Peritoneal fluid was collected with a 1-ml insulin syringe (BD Biosciences) and quantified by weight. Intestines (small intestine plus caecum plus colon) were harvested and immediately weighed. Harvested tissues were then dried uncovered overnight at 37 °C and weighed again. Fluid weight was calculated as wet weight minus dry weight.

Cell depletion and transfer. Neutrophils were depleted by injecting B6 mice with 200 μ g (intraperitoneal; 36 h before treatment) and 150 μ g (intravenous; 6 h before treatment) of RB6-8C5 antibody (anti-GR1; gift from D. Portnoy³²). Macrophages were depleted by injecting B6 mice intraperitoneally with 500 μ l and intravenously with 200 μ l of liposome-encapsulated clodronate (N.v.R) 48 h before FlaTox treatment. Clodronate liposomes were prepared as described previously³³. CD11b⁺ cells were depleted by injecting FVB:CD11b-DTR mice intraperitoneally with 25 ng g⁻¹ body weight of diphtheria toxin (Sigma-Aldrich) 24 h before FlaTox treatment.

For macrophage cell transfer (Fig. 2g), peritoneal cells were lavaged from naive or thioglycollate-treated mice and BMDMs were derived in culture. About 10⁷ macrophages (estimated by counting large cells in lavage) were injected intraperitoneally into host mice in 500 μ l of PBS. After 2 h, mice were injected with FlaTox and monitored for changes in rectal temperature. For spleen transfer, a single-cell suspension was generated by pushing spleen through mesh followed by osmotic lysis of red blood cells. Bone marrow cells were collected from femurs and tibias, followed by osmotic lysis of red blood cells. Entire spleen or bone marrow from one mouse was transferred into one host by intraperitoneal injection and analysed as above.

Eicosanoid analysis. For *ex vivo* analysis, total resident peritoneal cells were lavaged from euthanized mice and macrophage numbers estimated by counting only large cells. Macrophages (10^6 – 2×10^6 ; $(2$ – $4) \times 10^6$ total cells) were aliquoted to 1.5-ml Eppendorf tubes. Toxin treatments were performed in 1-ml prewarmed Dulbecco's PBS with Ca²⁺ and Mg²⁺ (Invitrogen). After 30 min, cells were pelleted by centrifugation and the supernatant was immediately transferred to 2 ml of cold methanol for storage at –80 °C. For *in vivo* analysis, mice were injected intraperitoneally with FlaTox. After 20 min, mice were euthanized and the peritoneum was lavaged with 1 ml of cold PBS, which was immediately transferred to 2 ml of cold methanol for storage at –80 °C. Before analysis, 400 pg each of the deuterated internal standards prostaglandin E₂ (PGE₂-d₄), 15(S)-hydroxyeicosatetraenoic acid (15(S)-HETE-d₈) and LTB₄-d₄ were added

to each sample to calculate the recovery of different classes of oxygenated fatty acid. Lipid autacoids were extracted by solid phase with SampliQ ODS-C₁₈ cartridges (Agilent Technologies). Eicosanoids and docosanoids were identified and quantified by LC–MS/MS-based lipidomics^{34–37}. In brief, we analysed extracted samples by a triple-quadrupole linear ion-trap LC–MS/MS system (MDS SCIEX 3200 QTRAP) equipped with a Kinetex C₁₈ mini-bore column. The mobile phase was a gradient of A (water/acetonitrile/acetic acid (72:28:0.01 by vol.)) and B (propan-2-ol/acetonitrile (60:40, v/v)) with a flow rate of 450 μ l min⁻¹. MS/MS analyses were performed in negative-ion mode, and prominent fatty acid metabolites were quantified by multiple reaction monitoring (MRM mode) using established transitions^{34–37} for PGE₂/PGD₂ (351→271, 351→189 m/z), TXB₂ (369→169 m/z), PGF_{2 α} (353→193 m/z), 5-HETE (319→115 m/z), 12-HETE (319→179 m/z), 15-HETE (319→175 m/z), 5,12-DiHETE/LTB₄ (335→195 m/z), LXA₄ (351→115 m/z), PGE₂-d₄ (355→275 m/z), LTB₄-d₄ (339→197 m/z), and 15-HETE-d₈ (327→182 m/z). Calibration curves (1–1,000 pg) and specific LC retention times for each compound were established with synthetic standards (Cayman Chemical). Structures were confirmed for selected autacoids by MS/MS analyses using enhanced product ion mode with appropriate selection of the parent ion in quadrupole 1.

For enzyme immunoassay (EIA) of LTB₄, PGE₂ and cysteinyl LTs, total resident peritoneal cells were collected by lavage as above. Macrophages (2×10^5) were seeded into 96-well plates and incubated for 4 h in RPMI buffer containing 10% FBS, 100 U ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin, 2 mM L-glutamine. Before assay, cells were washed once with PBS to select for adherent macrophages. Cells were treated for 30 min in 100 μ l of PBS with Ca²⁺ and Mg²⁺. PGE₂, LTB₄ and cysteinyl LTs in supernatants were measured by EIA (Cayman Chemical).

Salmonella infections. Resident peritoneal macrophages were selected from total peritoneal lavage by plating overnight on Petri dishes followed by rinsing with PBS before replating into 96-well plates (2×10^5 cells per well). *Salmonella enterica* serovar Typhimurium strain LT2 was grown overnight in 5 ml of Luria-Bertani medium in a shaking incubator at 37 °C. The next morning, the cultures were diluted 1:66 in Luria-Bertani medium and grown for 3 h (standing culture, 37 °C). Bacteria were added to cells at a multiplicity of infection of 5 in PBS with Ca²⁺ and Mg²⁺, followed by centrifugation at 400g for 10 min. After 3 h, eicosanoid production and cell lysis were measured by EIA and LDH assay, respectively, as described above.

Flow cytometry. Leukocytes were collected from the peritoneal cavity by lavage with 7 ml of PBS, from spleen passed through a nylon mesh (BD Falcon) to establish a single-cell suspension, or from the lamina propria as described previously³⁸. Cells were stained with anti-F4-80-APC (BM8; 1:200 dilution; eBiosciences), anti-CD11b-PE (M1/70; 1:400 dilution; eBiosciences) and anti-CD11c-PE-Cy7 (N418; 1:100 dilution; eBiosciences) and analysed with standard flow cytometry protocols. F4-80/CD11b^{hi} cells were isolated by fluorescence-activated cell sorting.

Quantitative RT-PCR. CD11b/F4-80^{hi} cells were sorted from total peritoneal cells lavaged from naive or thioglycollate-injected (2 ml injected intraperitoneally 4 days in advance) mice. Bone-marrow-derived macrophages are more than 95% CD11b/F4-80^{hi} and were used without sorting. RNA was isolated with the RNeasy kit (Qiagen) in accordance with the manufacturer's protocol. RNA samples were treated with RQ1 DNase (Promega) before reverse transcription with Superscript III (Invitrogen). Complementary cDNA reactions were primed with poly(dT) for the measurement of mature transcripts. Quantitative PCR was performed as described³⁹ using the Step One Plus RT PCR System (Applied Biosystems) with Platinum Taq DNA polymerase (Invitrogen) and EvaGreen (Biotium). Transcript levels were normalized to *Rps17*. The following primers were used in this study: *Rps17* sense, 5'-CGCCATTATCCCCAGCAAG-3'; *Rps17* antisense, 5'-TGTCGGATCCACCTCAATG-3'; *Ptgs1*(Cox1) sense, 5'-ATGAGTCGAAGGAGTCTCTCG-3'; *Ptgs1*(Cox1) antisense, 5'-GCACGGATAGTAACAACAGGGA-3'; *Alox5* sense, 5'-ACTACATCTACCTCAGCCTCATT-3'; *Alox5* antisense, 5'-GGTGACATCGTAGGAGTCCAC-3'; *Alox12/15* sense, 5'-GGTCCAA CAACGAGGTCTAC-3'; *Alox12/15* antisense, 5'-AGGTATTCTGACACA TCCACCTT-3'.

Western blotting. Secreted proteins from FlaTox-treated cells were collected and probed with anti-CASP1 p10 (sc-514; Santa Cruz) as described previously⁴. Cell lysates from 2×10^6 resident peritoneal cells were probed with anti-cPLA₂ (2832; Cell Signaling), anti-phospho-cPLA₂ (2831; Cell Signaling) and anti- β -actin (sc-47778; Santa Cruz).

Calcium flux. Resident peritoneal macrophages were selected from total peritoneal lavage by plating overnight on untreated Petri dishes followed by one rinse with PBS and replating onto eight-chamber glass slides (Thermo Scientific) coated with poly(D-lysine) (Sigma-Aldrich). After incubation overnight, cells were incubated for 45 min at 37 °C with 2.5 mM Fluo-4 (Invitrogen) plus 0.02% pluronic (Invitrogen) in Ringer's buffer containing 2 mM Ca²⁺. Cells were washed twice and incubated for 45 min in Ringer's buffer at 37 °C before transfer to microscope

for imaging. Cells were maintained at 37 °C with CO₂ throughout imaging. Fluorescent (470 nm/520 nm) and differential interference contrast images were collected every 10 s on a Nikon Eclipse TE 2000-E microscope and analysed with NIS Elements AR 3.2 software. At least 30 cells were analysed for each replicate. Onset of cell lysis was calculated as the earliest time point at which membrane blebbing was visible by differential interference contrast imaging.

30. Mariathasan, S. *et al.* Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. *Nature* **430**, 213–218 (2004).
31. Li, P. *et al.* Mice deficient in IL-1 β -converting enzyme are defective in production of mature IL-1 β and resistant to endotoxic shock. *Cell* **80**, 401–411 (1995).
32. Glomski, I. J., Decatur, A. L. & Portnoy, D. A. *Listeria monocytogenes* mutants that fail to compartmentalize listeriolysin O activity are cytotoxic, avirulent, and unable to evade host extracellular defenses. *Infect. Immun.* **71**, 6754–6765 (2003).
33. Van Rooijen, N. & Sanders, A. Liposome mediated depletion of macrophages: mechanism of action, preparation of liposomes and applications. *J. Immunol. Methods* **174**, 83–93 (1994).
34. Sapieha, P. *et al.* 5-Lipoxygenase metabolite 4-HDHA is a mediator of the antiangiogenic effect of ω -3 polyunsaturated fatty acids. *Sci. Transl. Med.* **3**, 69ra12 (2011).
35. Liclican, E. L., Nguyen, V., Sullivan, A. B. & Gronert, K. Selective activation of the prostaglandin E₂ circuit in chronic injury-induced pathologic angiogenesis. *Invest. Ophthalmol. Vis. Sci.* **51**, 6311–6320 (2010).
36. Leedom, A. J., Sullivan, A. B., Dong, B., Lau, D. & Gronert, K. Endogenous LXA₄ circuits are determinants of pathological angiogenesis in response to chronic injury. *Am. J. Pathol.* **176**, 74–84 (2010).
37. Hassan, I. R. & Gronert, K. Acute changes in dietary ω -3 and ω -6 polyunsaturated fatty acids have a pronounced impact on survival following ischemic renal injury and formation of renoprotective docosahexaenoic acid-derived protectin D1. *J. Immunol.* **182**, 3223–3232 (2009).
38. Annacker, O. *et al.* Essential role for CD103 in the T cell-mediated regulation of experimental colitis. *J. Exp. Med.* **202**, 1051–1061 (2005).
39. Monroe, K. M., McWhirter, S. M. & Vance, R. E. Identification of host cytosolic sensors and bacterial factors regulating the type I interferon response to *Legionella pneumophila*. *PLoS Pathog.* **5**, e1000665 (2009).

Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus

Augustin de Vanssay^{1†}, Anne-Laure Bougé^{2†}, Antoine Boivin¹, Catherine Hermant¹, Laure Teyssset¹, Valérie Delmarre¹, Christophe Antoniewski^{2‡} & Stéphane Ronsseray¹

A paramutation is an epigenetic interaction between two alleles of a locus, through which one allele induces a heritable modification in the other allele without modifying the DNA sequence^{1,2}. The paramutated allele itself becomes paramutagenic, that is, capable of epigenetically converting a new paramutable allele. Here we describe a case of paramutation in animals showing long-term transmission over generations. We previously characterized a homology-dependent silencing mechanism referred to as the *trans*-silencing effect (TSE), involved in *P*-transposable-element repression in the germ line^{3–5}. We now show that clusters of *P*-element-derived transgenes that induce strong TSE^{6,7} can convert other homologous transgene clusters incapable of TSE into strong silencers, which transmit the acquired silencing capacity through 50 generations. The paramutation occurs without any need for chromosome pairing between the paramutagenic and the paramutated loci, and is mediated by maternal inheritance of cytoplasm carrying Piwi-interacting RNAs (piRNAs) homologous to the transgenes. The repression capacity of the paramutated locus is abolished by a loss-of-function mutation of the *aubergine* gene involved in piRNA biogenesis, but not by a loss-of-function mutation of the *Dicer-2* gene involved in siRNA production. The paramutated cluster, previously producing barely detectable levels of piRNAs, is converted into a stable, strong piRNA-producing locus by the paramutation and becomes fully paramutagenic itself. Our work provides a genetic model for the emergence of piRNA loci, as well as for RNA-mediated *trans*-generational repression of transposable elements.

Paramutations have been well described in plants^{1,2,8–12}. The best characterized is the *b1* paramutation in maize, which involves a small RNA silencing pathway^{13–15}, changes in DNA methylation levels and chromatin modifications¹⁶, and shows full penetrance and stability across generations. Paramutation-like phenomena involving microRNAs have been described in mice^{17,18}. However, long-term inheritance of a paramutation through generations has not been reported so far in animals.

In *Drosophila melanogaster*, transposition of *P* elements causes hybrid dysgenesis, a syndrome of genetic abnormalities including a high mutation rate, chromosome rearrangements and sterility^{19,20}. In natural populations, telomeric *P* elements inserted in heterochromatic telomere-associated sequences (TAS) are master sites for establishing *P*-element repression in the germ line^{21–23}. In laboratory lines (for example, *P-1152*), *P-lacZ* transgenes inserted in TAS mimics telomeric *P* elements by repressing germline expression of reporter transgenes inserted at distant euchromatic sites, through a homology-dependent silencing mechanism, TSE^{3–5,24}. TSE is strongly sensitive to mutations affecting the piRNA pathway^{5,25}. Its establishment involves both genetic and epigenetic components: a chromosomal copy of the telomeric silencer transgene must be either paternally or maternally inherited, and a cytoplasmic component containing small RNAs homologous to the transgene must be maternally inherited^{4,5}. In

addition to telomeric loci, we found that *T-1*, a tandem repeat cluster of *P-lacZ* transgenes inserted in the middle of chromosome arm 2R (50C), can also trigger a strong TSE⁷. *T-1* and other *P-lacZ* clusters inserted at the same locus (Supplementary Fig. 1) induce ectopic heterochromatin and show variegation of the *white* gene marker in the eye, a phenomenon termed repeat-induced gene silencing^{6,26}. However, *T-1* triggers strong silencing of various TSE reporter transgenes in the germ line⁷, whereas the other transgene clusters at this locus, including *BX2*, which contains the same number of transgene repeats as *T-1*, did not induce detectable TSE (Supplementary Table 1).

The epigenetic properties of *T-1* were analysed together with those of the *P-1152* telomeric silencer and the *BX2* cluster as controls. *T-1* and *P-1152* showed typical maternal transmission of TSE: strong repression occurred in the germ line of progeny when the silencer was maternally inherited (Fig. 1a), whereas weak or null repression was detected when the silencer was paternally inherited (Fig. 1b). *BX2* showed no repression capacity in these crosses. To analyse the relationship between TSE and piRNAs, we sequenced 19–29-nucleotide RNAs from ovaries of *T-1*, *P-1152* or *BX2* females (Supplementary Table 2). Abundant small RNAs matched the *T-1* sequences in the library from hemizygous females having inherited the *T-1* locus maternally (Fig. 1c), but not paternally (Fig. 1d). Among these species, the 23–28-nucleotide RNAs showed the typical ‘ping-pong’ signature of piRNA biogenesis²⁷, including a bias for a 5′ U (1U) and a strong tendency to form sense–antisense pairs with complementarity over their first ten nucleotides (Supplementary Fig. 2). In addition to piRNAs, short interfering RNAs (siRNAs) have been shown to be produced by previously characterized piRNA loci²⁸. Similarly, *T-1* produced a significant fraction of 21-nucleotide RNAs (Fig. 1c) that do not show the ping-pong signature of piRNAs and probably correspond to siRNAs (Supplementary Fig. 3a). In agreement with a previous report²⁹, small RNAs with similar features were produced by *P-1152* in hemizygous females having inherited the *P-1152* locus maternally (Fig. 1f). Homozygous *P-1152* females produced about twice as many piRNAs as these hemizygous females (Supplementary Fig. 4). Finally, only a very low level of small RNAs was produced that matched *BX2* in hemizygous females from the *BX2* line (Fig. 1e). Hence, maternal inheritance of *T-1*, as well as *P-1152*, is associated with both the production of piRNAs derived from these loci and the capacity of these loci to mediate TSE, thereby linking silencing and piRNAs in this system. We next tested epigenetic interactions between the *P-1152* telomeric silencer and *T-1*, and found that chromosomal and maternally transmitted components of *T-1* and *P-1152* can complement each other to induce TSE (Supplementary Fig. 5), consistent with the presence of piRNAs matching *P-lacZ* sequences in ovaries of both *T-1* and *P-1152* females.

To investigate possible transfer of epigenetic information between *T-1* and the inactive *BX2* locus, we crossed hemizygous *T-1* females

¹Laboratoire Biologie du Développement, UMR7622, CNRS-Université Pierre et Marie Curie, 9 quai Saint Bernard, 75005 Paris, France. ²Drosophila Genetics and Epigenetics, CNRS URA2578 - Institut Pasteur, 25 rue du Dr Roux, 75015 Paris, France. [†]Present addresses: Institut Jacques Monod, CNRS, UMR 7592, Université Diderot, Sorbonne Paris Cité, F-75205 Paris, France (A.d.V.); Drosophila Normal and Pathological Neurobiology, INSERM U661 - Institut de Génétique Fonctionnelle, 141 rue de la Cardonille, 34094 Montpellier, France (A.-L.B.); Drosophila Genetics and Epigenetics, Laboratoire Biologie du Développement, UMR7622, CNRS-Université Pierre et Marie Curie, 9 quai Saint Bernard, 75005 Paris, France (C.A.).

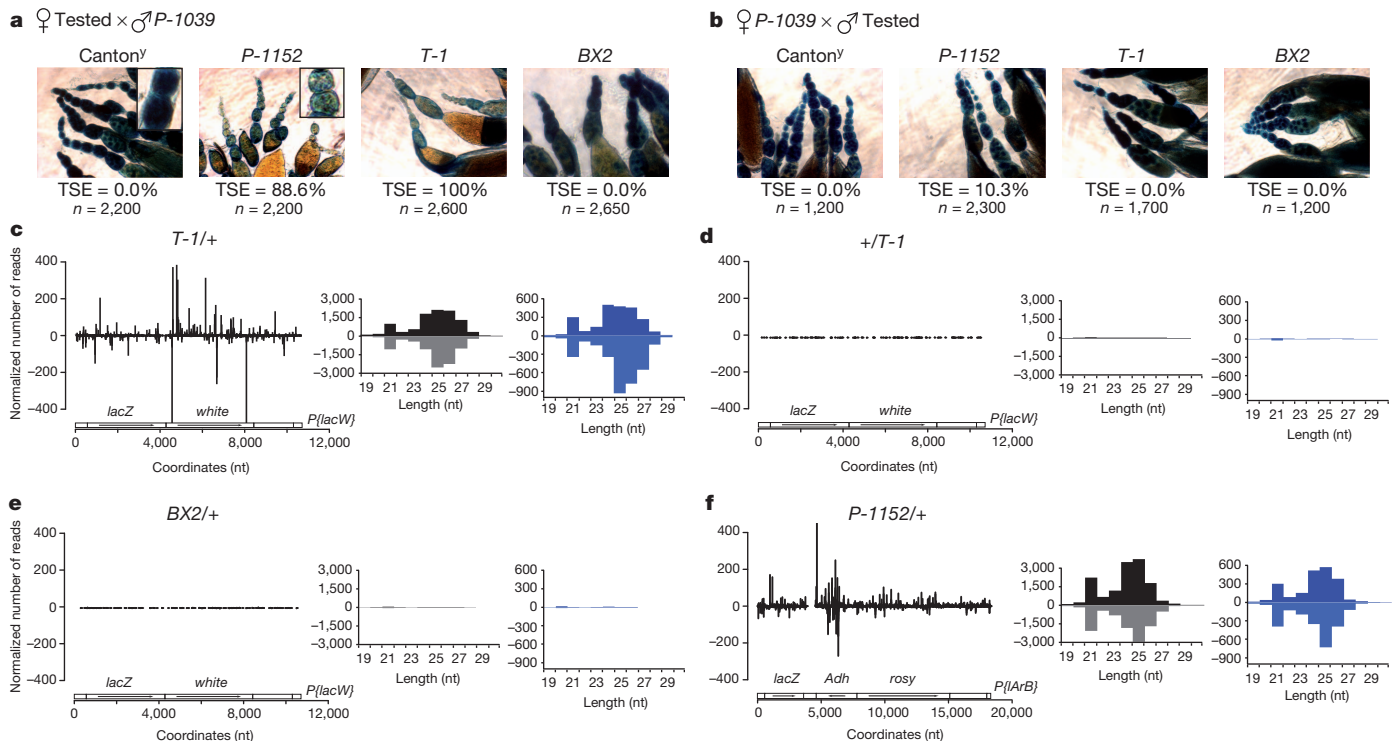


Figure 1 | Maternal inheritance of *P-1152* and *T-1* repression capacities correlates with the presence of *T-1*- or *P-1152*-derived piRNAs in ovaries of female progeny. **a, b**, Maternal (**a**) and paternal (**b**) inheritance of TSE mediated by *P-1152*, *T-1* or *BX2* was tested using the *P-1039* TSE reporter transgene. *lacZ* staining of ovaries of G₁ females from the indicated crosses was performed, and TSE was expressed as the percentage of repressed egg chambers among the total number (*n*) of egg chambers analysed. Female and male Canton^y flies are devoid of any transgene and were used as controls. Note that *lacZ* staining of follicle cells surrounding egg chambers (shown at higher

magnification in insets) is observed in all ovaries because TSE only occurs in the germ line⁴. Original magnification, ×20. **c–f**, Deep sequencing of small RNAs from ovaries of the indicated genotypes in which the maternally inherited allele is always indicated first. Plots show the abundance of 19–30-nucleotide (nt) small RNAs matching *P[lacW]* (**c–e**) or *P[lArB]* (**f**). Histograms show the length distributions of small RNAs matching *P[lacW]* or *P[lArB]* (dark bars), or only the *lacZ* sequence in these elements (blue bars). Positive and negative values correspond to sense and antisense reads, respectively.

with hemizygous *BX2* males, and recovered female progeny that had not inherited the *T-1* locus and carrying a paternally inherited *BX2* locus (Fig. 2). These females showed marked silencing of the TSE reporter transgene, indicating that the cytoplasm of *T-1* oocytes can

confer new silencing capacities to the inactive allele of the *BX2* locus. This *de novo* silencing allele will be hereafter referred to as *BX2** to differentiate it from the initial *BX2* allele never having been exposed to a *T-1* cytoplasm.

A *BX2** line was established and analysed in successive generations (Fig. 3a). Notably, second generation (G₂) *BX2** females from test crosses with males carrying a TSE reporter transgene still showed a complete TSE (Fig. 3b). This capacity to mediate TSE was fully maintained over 25 generations of the *BX2** line (TSE = 100%, *n* = 4,600). TSE remained very strong between G₃₂ and G₅₅ (99.4%, *n* = 22,700) showing a reversion rate less than 0.5% per generation at 25 °C (Supplementary Discussion). We conclude that maternally inherited factors from the *T-1* strain stably paramutated the *BX2* locus.

In contrast to *BX2* females, ovaries of G₂ *BX2** females contained abundant small RNAs matching the *BX2* sequence (Fig. 3c and Supplementary Table 2) with a profile similar to the one observed in *T-1* females (see Fig. 1c). The size distribution of these small RNAs showed a large peak corresponding to 23–28-nucleotide small RNAs with the piRNA ping-pong signature (Supplementary Fig. 2), as well as a discrete peak corresponding to a 21-nucleotide siRNA-like species of RNAs. Therefore, the acquired capacity of the *BX2** allele to mediate TSE correlates with the *de novo* production of *lacZ*-derived small RNAs from this locus. Finally, *BX2**-derived small RNAs were continuously produced in ovaries over at least 42 generations of a *BX2** line (Fig. 3d and Supplementary Figs 2 and 3). Together, these data indicate that the *BX2** paramutation is associated with stable production of high levels of small RNAs from the *BX2* locus in ovaries.

We next tested whether the paramutated *BX2** allele is paramutagenic. We crossed hemizygous *BX2** females with hemizygous naive *BX2* males and recovered female progeny having inherited the

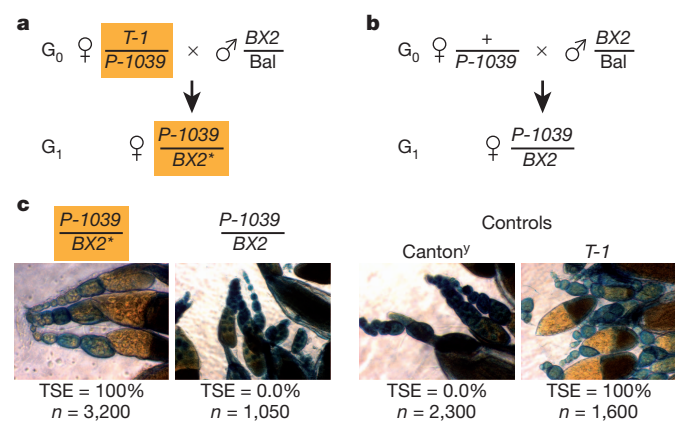


Figure 2 | Epigenetic induction of *BX2* by *T-1*. **a**, *T-1* females carrying the TSE reporter transgene *P-1039* were crossed to *BX2* males carrying a balancer chromosome (Bal). *BX2** female progeny having inherited cytoplasm from *T-1* mothers (orange background) and a *BX2* chromosome from fathers were stained for *lacZ*. **b**, Females carrying only the TSE reporter *P-1039* were crossed to *BX2* males. Female progeny from this cross were stained for *lacZ*. **c**, *P-1039*/*BX2** female progeny from the cross in **a** showed complete TSE, which was scored as indicated in Fig. 1. *P-1039*/*BX2* female progeny from the cross in **b** did not show TSE. Controls correspond to crosses between Canton^y (devoid of any transgene) or *T-1* females with *P-1039* males, which resulted in progeny showing null and complete TSE, respectively. Original magnification, ×20.

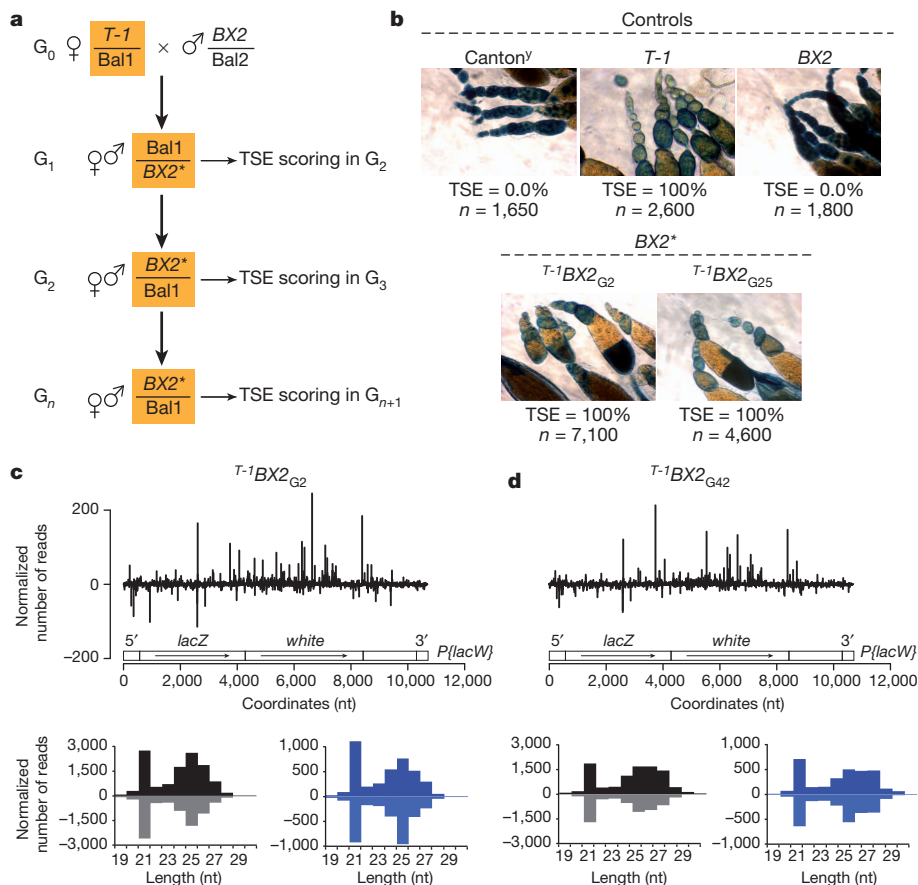


Figure 3 | *BX2** paramutation occurs and is associated to the production of small RNAs by the *BX2* cluster. **a**, *BX2** lines were established as indicated. Bal1 and Bal2 are balancer chromosomes carrying distinct phenotypic markers. *BX2** siblings were crossed at each generation to perpetuate the *BX2** line. In addition, *BX2** females were crossed at various generations (G_n) to males carrying the *P-1039* reporter, to score the TSE of *BX2** in the G_{n+1} female progeny. **b**, TSE in *BX2** females from generations G_2 and G_{25} , and in progeny of crosses from Canton^y, *T-1* and *BX2* females with *P-1039* males as controls. *T-1BX2*_{G25} indicates that *BX2* females inherited cytoplasm from *T-1* females 25 generations before the present cross. TSE was scored as indicated in Fig. 1. Original magnification, $\times 20$. **c**, **d**, Abundance (top) and length distribution (dark histograms) of 19–30-nucleotide small RNAs matching the *P[lacW]* transgene in ovaries from hemizygous *BX2** females from generation G_2 (**c**) and G_{42} (**d**). Length distributions of the subsets of small RNAs only matching *lacZ* are shown as blue histograms. Positive and negative values correspond to sense and antisense reads, respectively.

cytoplasm of *BX2** mothers and the *BX2* locus from fathers (Fig. 4a). This *BX2* allele was then assessed in generation G_2 for its capacity to silence a TSE reporter transgene in the germline. Notably, we observed a complete TSE (Fig. 4a), indicating that the paternally inherited *BX2* allele was paramutated through maternal inheritance of *BX2** cytoplasm. This newly paramutated *BX2* allele, which corresponds to a second-order paramutation, will be hereafter referred to as *BX2*^{*2}. A *BX2*^{*2} line was established and showed stable TSE over 36 generations (Fig. 4a). Moreover, this line retained the capacity to produce large amounts of *BX2*^{*2}-derived small RNAs after 36 generations (Fig. 4b). Following an identical mating scheme, *BX2*^{*2} females were able to paramutate a paternally inherited *BX2* locus, generating a third-order *BX2*^{*3} paramutated allele that showed full TSE capacity over 10 generations. Applying this procedure recurrently, we generated a

fifth-order paramutated *BX2*^{*5} allele that showed full TSE capacity (Supplementary Fig. 6). In conclusion, the conversion of *BX2* to *BX2*^{*} by *T-1* maternal cytoplasm has all the properties of a paramutation, because it is stable over generations and the paramutated allele shows secondary paramutagenicity.

Interestingly, *T-1* also fully paramutated C2, another seven-copy transgene inserted at the same location (Supplementary Fig. 1), whereas lower-copy-number transgenes at this location were paramutated only transiently (Supplementary Table 3). A similar unstable paramutation interaction was also observed between the non-allelic *P-1152* and *BX2* loci (Supplementary Fig. 7).

As paramutation in this system is correlated with the production of *BX2*^{*}-derived piRNAs and siRNAs, we investigated the effect of *aubergine* and *Dicer-2* loss of function on a paramutated *BX2* cluster.

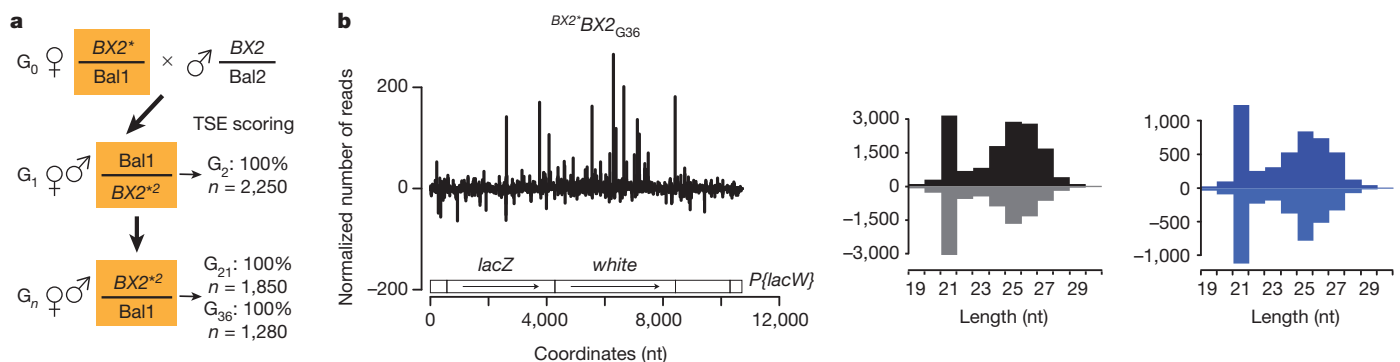


Figure 4 | Paramutated *BX2** is paramutagenic. **a**, *BX2** females were crossed with *BX2* males and a *BX2*^{*2} line (second-order paramutation) was established as indicated. Bal1 and Bal2 are balancer chromosomes. *BX2*^{*2} siblings were crossed at various generations to perpetuate the *BX2*^{*2} line. In addition, *BX2*^{*2} females were crossed at each generation (G_n) with males carrying the *P-1039* reporter transgene to score the TSE of *BX2*^{*2} in the G_{n+1}

female progeny. **b**, Abundance (graph on the left) and length distribution (black histogram in the middle) of 19–30-nucleotide small RNAs matching the *P[lacW]* transgene in ovaries from hemizygous *BX2*^{*2} females from generation G_{36} . Length distribution of the subset of small RNAs only matching *lacZ* is shown in the blue histogram on the right.

The silencing capacity of the *BX2*² cluster was completely abolished in homozygous *aubergine* mutants, whereas strong silencing still took place in *Dicer-2* homozygous mutants (Supplementary Fig. 8). Moreover, the *BX2*² locus still showed full repression capacity after four generations in a *Dicer-2* homozygous mutant context. Hence, the *BX2*² silencing activity requires piRNAs, whereas neither *BX2*² activity nor inheritance rely on siRNAs. In maize, paramutation can be induced by a non-allelic transgene producing *b1*-repeat double-stranded RNA (dsRNA) and siRNAs¹⁵ and epigenetic inheritance of the *Kit*^{tm1Alf} mutant allele in mice seems to result from paternal as well as maternal transmission of small RNAs¹⁷. These data indicate that paramutations may in some instances involve small RNAs without interactions between alleles at the DNA or chromatin levels. Our findings that, in *Drosophila*, the *BX2* paramutation is triggered by cytoplasmic inheritance strongly support this view.

Finally, we investigated the effect of the paramutation on transcription of the *BX2* locus by quantitative polymerase chain reaction with reverse transcription (RT-qPCR). *BX2* and *BX2*² showed similar steady-state levels of both sense and antisense transcripts (Supplementary Fig. 9). This observation suggests that paramutation, rather than increasing the pool of piRNA precursor transcripts, activates their downstream processing into piRNAs. Thus, the maternally transmitted piRNAs could trigger production of primary piRNAs and/or ping-pong amplification of secondary piRNAs in the nuage. As paramutation is accompanied by *de novo* production of high levels of piRNA, it provides an invaluable model to determine the molecular events involved in the genesis of piRNA loci.

METHODS SUMMARY

All crosses were performed at 25 °C. *lacZ* expression assays were carried out using X-gal overnight staining³⁰. The *P-lacZ-white* construct (named *P[lacW]*) contains the *P-lacZ* translational fusion and is marked by the mini-*white* gene (Supplementary Fig. 1 and Supplementary Table 4). Small RNA libraries from hand-dissected ovaries were prepared using the Illumina kit and sequenced using an Illumina Genome Analyzer II or an Illumina HiSeq-2000, following the manufacturer's instructions. For library comparisons, read counts were normalized to the total number of small RNAs that matched the *D. melanogaster* genome and did not correspond to abundant cellular RNAs (ribosomal RNA, transfer RNA and small nucleolar RNAs). Overlap signatures were computed for each sequence data set by collecting the appropriate RNA reads matching *P* transgenes and calculating overlap frequencies with RNA reads on the opposite strand.

Full Methods and any associated references are available in the online version of the paper.

Received 6 June 2011; accepted 16 July 2012.

Published online 26 August 2012.

- Brink, R. A. A genetic change associated with the R locus in maize which is directed and potentially reversible. *Genetics* **41**, 872–889 (1956).
- Coe, E. H. Jr. A regular and continuing conversion-type phenomenon at the B locus in maize. *Proc. Natl Acad. Sci. USA* **45**, 828–832 (1959).
- Roche, S. E. & Rio, D. C. Trans-silencing by *P* elements inserted in subtelomeric heterochromatin involves the *Drosophila* Polycomb group gene, *Enhancer of zeste*. *Genetics* **149**, 1839–1855 (1998).
- Josse, T. et al. Telomeric trans-silencing in *Drosophila melanogaster*: tissue specificity, development and functional interactions between non-homologous telomeres. *PLoS ONE* **3**, e3249 (2008).
- Josse, T. et al. Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. *PLoS Genet.* **3**, 1633–1643 (2007).
- Dorer, D. R. & Henikoff, S. Transgene repeat arrays interact with distant heterochromatin and cause silencing in *cis* and *trans*. *Genetics* **147**, 1181–1190 (1997).
- Ronsseay, S., Boivin, A. & Anxolabehere, D. P-element repression in *Drosophila melanogaster* by variegating clusters of *P-lacZ-white* transgenes. *Genetics* **159**, 1631–1642 (2001).
- Chandler, V. L. Paramutation: from maize to mice. *Cell* **128**, 641–645 (2007).
- Hollick, J. B., Patterson, G. I., Coe, E. H., Cone, K. C. & Chandler, V. L. Allelic interactions heritably alter the activity of a metastable maize *pl* allele. *Genetics* **141**, 709–719 (1995).

- Pilu, R. et al. A paramutation phenomenon is involved in the genetics of maize *low phytic acid1-241* (*lpa1-241*) trait. *Heredity* **102**, 236–245 (2009).
- Sidorenko, L. V. & Peterson, T. Transgene-induced silencing identifies sequences involved in the establishment of paramutation of the maize *p1* gene. *Plant Cell* **13**, 319–335 (2001).
- Stam, M. Paramutation: a heritable change in gene expression by allelic interactions in *trans*. *Molecular Plant* **2**, 578–588 (2009).
- Alleman, M. et al. An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature* **442**, 295–298 (2006).
- Dorweiler, J. E. et al. mediator of paramutation1 is required for establishment and maintenance of paramutation at multiple maize loci. *Plant Cell* **12**, 2101–2118 (2000).
- Arteaga-Vazquez, M. et al. RNA-mediated trans-communication can establish paramutation at the *b1* locus in maize. *Proc. Natl Acad. Sci. USA* **107**, 12986–12991 (2010).
- Stam, M., Bebele, C., Dorweiler, J. E. & Chandler, V. L. Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation. *Genes Dev.* **16**, 1906–1918 (2002).
- Rassoulzadegan, M. et al. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469–474 (2006).
- Grandjean, V. et al. The *miR-124-Sox9* paramutation: RNA-mediated epigenetic control of embryonic and adult growth. *Development* **136**, 3647–3655 (2009).
- Kidwell, M. G., Kidwell, J. F. & Sved, J. A. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility, and male recombination. *Genetics* **86**, 813–833 (1977).
- Engels, W. R. in *P Elements in Drosophila* (eds Berg, D. E. & Howe, M. M.) (American Society for Microbiology, 1989).
- Ronsseay, S., Lehmann, M., Nouaud, D. & Anxolabehere, D. The regulatory properties of autonomous subtelomeric *P* elements are sensitive to a suppressor of variegation in *Drosophila melanogaster*. *Genetics* **143**, 1663–1674 (1996).
- Marin, L. et al. P-element repression in *Drosophila melanogaster* by a naturally occurring defective telomeric P copy. *Genetics* **155**, 1841–1854 (2000).
- Stuart, J. R. et al. Telomeric P elements associated with cytotypic regulation of the P transposon family in *Drosophila melanogaster*. *Genetics* **162**, 1641–1654 (2002).
- Poyhonen, M. et al. Homology-dependent silencing by an exogenous sequence in the *Drosophila* germline. *G3 (Bethesda)* **2**, 331–338 (2012).
- Todeschini, A. L., Teyssier, L., Delmarre, V. & Ronsseay, S. The epigenetic trans-silencing effect in *Drosophila* involves maternally-transmitted small RNAs whose production depends on the piRNA pathway and HP1. *PLoS ONE* **5**, e11032 (2010).
- Dorer, D. R. & Henikoff, S. Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. *Cell* **77**, 993–1002 (1994).
- Brennecke, J. et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
- Czech, B. et al. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453**, 798–802 (2008).
- Muerdter, F. et al. Production of artificial piRNAs in flies and mice. *RNA* **18**, 42–52 (2011).
- Lemaître, B., Ronsseay, S. & Coen, D. Maternal repression of the P element promoter in the germline of *Drosophila melanogaster*: a model for the P cytotypic. *Genetics* **135**, 149–160 (1993).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank O. Sismeiro, J.-Y. Copée, E. Mouchel-Vielh, V. Ribeiro, C. Pappatoc and P. Graça for technical assistance, D. Dorer, S. Henikoff and the Bloomington Stock Center for providing stocks, and flybase.org for providing databases. We thank T. Josse for preliminary experiments. We thank J.-R. Huynh, V. Colot, N. Randsholt, A.-M. Pret, C. Carré and F. Peronnet for critical reading of the manuscript. S.R. thanks D. Anxolabehere and M. Lehmann for previous help. This work was supported by fellowships from the Ministère de l'Enseignement Supérieur et de la Recherche to A.d.V. and C.H., from the Fondation pour la Recherche Médicale to A.d.V., from the Association Nationale de la Recherche (ANR) to A.-L.B., and by grants from the Association pour la Recherche contre le Cancer to S.R. and from the ANR (project “Nuclear endosRNAs”) to C.A.

Author Contributions Genetic experiments were conceived by A.d.V., A.B. and S.R., and performed by A.d.V., A.B., C.H., V.D., L.T. and S.R. L.T. conceived and performed molecular mapping of the clusters and Southern blot analysis. Deep-sequencing analysis was conceived by A.d.V., A.-L.B., S.R. and C.A., and performed by A.d.V. and A.-L.B. Bioinformatic analysis was conceived and performed by C.A. RT-qPCR was conceived and performed by A.B. S.R., A.d.V., A.B. and C.A. wrote the paper and all authors discussed the results.

Author Information Small RNA sequences have been deposited at the National Center for Biotechnology Information under accession SRP012172. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.R. (stephane.ronsseay@upmc.fr) or C.A. (christophe.antoniowski@upmc.fr).

METHODS

Experimental conditions. All crosses were performed at 25 °C and involved 3–5 couples in most cases. *lacZ* expression assays were carried out using X-gal overnight staining as described previously³⁰, except that ovaries were fixed for 6 min. **Transgenes and strains.** *P-lacZ* fusion enhancer trap transgenes *P-1152*, *BQ16*, *BC69* and *P-1039* all contain an in-frame translational fusion of the *Escherichia coli lacZ* gene to the second exon of the *P* transposase gene and a *rosy* transformation marker³¹. The *P-1152* insertion (Supplementary Table 4) was mapped to the telomere of the X chromosome (cytological site 1A) and consists of two *P-lacZ* insertions in the same TAS unit and in the same orientation⁵. *P-1152* is homozygous, viable and fertile. *BQ16* is located at 64C in euchromatin of the third chromosome⁴ (Supplementary Table 4) and is homozygous, viable and fertile. *BC69* is inserted in chromosome 2 (Supplementary Table 4) in the first exon of the *vasa* gene and results in a *vasa* loss-of-function allele; consequently, it is homozygous, female and sterile. *P-1039* is located at 60B on the second chromosome (Supplementary Table 4) and is homozygous lethal. *P-1152* shows no *lacZ* expression in the ovary, *BQ16* and *BC69* are strongly expressed in the nurse cells and in the oocyte and *P-1039* shows strong *lacZ* staining in numerous tissues including the follicle cells, the nurse cells and the oocyte.

***P-lacZ* clusters.** Lines with different numbers of *P-lacZ-white* transgenes³² located at cytological site 50C on the second chromosome^{6,26} were used (Supplementary Table 4). The transgene(s) insertion site is located near the *mRPL53* gene, in an *Ago1* intron. This site is not a piRNA-producing locus, as observed for instance in the deep-sequencing data set from *P-1152* ovaries (data not shown). The *P-lacZ-white* construct contains the *P-lacZ* translational fusion and is marked by the mini-white gene (*P{lacW}*, FBtp0000204). *BX2* carries seven *P-lacZ* copies including at least one defective copy inserted in direct orientations. *T-1* derives from *BX2* following X-ray treatments (Supplementary Fig. 1). *T-1* has chromosomal rearrangements including translocations between the second and the third chromosomes. After overnight staining, weak *lacZ* expression is detected in the follicle cells of *BX2* and *T-1* female ovaries, presumably because of a position effect at 50C, but no staining is observed in the germ line (data not shown).

Lines carrying transgenes have *M* genetic backgrounds (devoid of *P* transposable elements), as do the multi-marked balancer stocks used in genetic experiments. The Canton^y and *w*¹¹¹⁸ lines were used as controls completely devoid of any *P* element or transgene. Crosses involving *P-1152* were performed with females carrying the telomeric transgenes in the homozygous state (except where indicated), whereas crosses performed with *BX2* or *T-1* were performed with females carrying the cluster in the heterozygous state (referred to as hemizygous in case of insertions) because of the sterility (*BX2*) and lethality (*T-1*) induced by transgene clusters.

Two strong hypomorphic mutant alleles of *aubergine* induced by EMS were used. Both of them are homozygous, female and sterile, and TSE was previously shown to be abolished by a heteroallelic combination of these alleles⁵. *aub*^{QC42} comes from the Bloomington Stock Center (stock no. 4968) and has not been characterized at the molecular level³³. *aub*^{N11} has a 154-bp deletion, resulting in a frameshift which is predicted to add 16 novel amino acids after residue 740 (refs 34, 35). *Dicer-2*^{L811fsX} is a loss-of-function allele induced by EMS that has a sequence variant at residue 811 resulting in a stop codon³⁶. It is homozygous, viable and fertile.

Quantification of TSE. When TSE is incomplete, variegation is observed because 'on/off' *lacZ* expression is seen between egg chambers: that is, egg chambers can show strong expression (dark blue) or no expression, but intermediate expression levels are rarely found. TSE was quantified as previously described⁵ by determining the percentage of egg chambers with no expression in the germ line.

Deep sequencing analyses. Small RNAs from hand-dissected ovaries were cloned using the DGE-Small RNA Sample Prep Kit and the Small RNA Sample Prep v.1.5 Conversion Kit from Illumina (libraries 1 to 5), following the manufacturer's instructions, or using the TruSeq (TM) SBS v.5 Kit at FASTERIS (<http://www.fasteris.com/>) (libraries 6 to 8). Libraries 1 to 5 were sequenced using an Illumina Genome Analyzer II and libraries 6 to 8 were sequenced using an Illumina HiSeq 2000. Sequence reads in fastq format were trimmed from the adaptor sequence 5'-TCGTATGCCCTCTTCTGCTTG-3' (libraries 1 to 5) or 5'-CTGTAGG CACCATCAAT-3' (libraries 6 to 8) and matched to the *D. melanogaster* genome release 5.43 using Bowtie³⁷, as well as to the sequences of the *P*-element constructs *P{larB}* (FlyBase accession FBtp0000160) and *P{lacW}* (FlyBase accession FBtp0000204). Only 19–30-nucleotide reads matching the reference sequences with 0 or 1 mismatch were retained for subsequent analysis. For global annotation of the libraries (Supplementary Table 2), we used release 5.43 of fasta reference files available in FlyBase, including transposon sequences (dmel-all-transposon_r5.43.fasta) and release 18 of miRNA sequences from miRBase (<http://www.mirbase.org>).

Sequence length distributions, small RNA mapping and frequency maps were generated using in-house Python scripts and R (<http://www.r-project.org/>) to

analyse Bowtie outputs. Scripts were integrated and run in a Galaxy instance hosted by the laboratory. The corresponding Mississippi suite of analysis workflows and codes is accessible from <http://www.drosophila.org> upon request. For library comparisons, read counts were normalized (Supplementary Table 2) to the total number of small RNAs that matched the *D. melanogaster* genome and did not correspond to abundant cellular RNAs (rRNA, tRNA and snoRNAs). For small RNA mapping, we matched each individual RNA sequence to *P{larB}* or *P{lacW}* and gave to each matched position a weight corresponding to the normalized occurrence of the sequence in the small RNA library. When RNA sequences matched *P{larB}* or *P{lacW}* repeatedly, the weight was divided by the number of hits to these *P*-element constructs.

Distributions of piRNA overlaps (ping-pong signatures) were computed by collecting, for each sequencing data set, all the 23–28-nucleotide RNA reads matching *P{larB}* or *P{lacW}* whose 5' ends overlapped with another 23–28-nucleotide RNA read on the opposite strand. Then, for each possible overlap of 1–28 nucleotides, the number of read pairs was counted. Distributions of siRNA overlaps were computed using a similar procedure, except that 20–22-nucleotide RNA reads were collected instead of the 23–28-nucleotide RNA reads. The distributions of piRNA/siRNA overlaps were computed by collecting separately the 20–22-nucleotide and 23–28-nucleotide RNA reads matching *P{larB}* or *P{lacW}*, and counting for each possible overlap of 1–22 nucleotides the number of read pairs across these two distinct read data sets. To plot the overlap signatures, a *z*-score was calculated by computing, for each overlap of 1 to *i* nucleotides, the number *O(i)* of read pairs and converting it using the formula $z(i) = (O(i) - \text{mean}(O)) / \text{standard deviation}(O)$.

RT-qPCR experiments. Total RNA was extracted (Qiagen kit) from ovaries dissected from 1A-6, *BX2* and *BX2** females and quantified (NanoDrop). Four to six biological replicates were made for each genotype. For each sample, 10 µg of RNA was treated with DNase (Fermentas). 1 µg of DNase-treated RNA was used for reverse transcription (Fermentas) using either no primer (control RT) or two primers simultaneously (specific RT): one specific to the *nanos* transcript used as the sample RNA quantification reference (5'-GGATTCGCCCTCTCTAAACC-3') and the second specific to a region of the *P{lacW}* transgene. *P{lacW}* RT primers were designed to be specific to the sense (s) or to the antisense (a) transcripts of five regions of the *P{lacW}* transgene: 5'*P*, 5'*lacZ*, 3'*lacZ*, 5'*white* and 3'*P*. Sequences are: a1 (5'-ATTCAAACCCACGGACAT-3'), a2 (5'-AGTACGAAATGCGTCGTTTAGAGC-3'), a3 (5'-GGGGAAAACCTTATTATCAGCCG-3'), a4 (5'-GCTGTTTGCTCTCTCTCTG-3'), s1 (5'-GTTTCCAGTCACGACGT-3'), s2 (5'-AATGCGCTCAGTCAAATTC-3'), s3 (5'-TATGGAACCGTCGATATTCAGCC-3'), s4 (5'-ATTTTGTGGGTCGCGAGTTC-3'), s5 (5'-TTAAGTGTATACTTCGGTAAGCTTCG-3'), s6 (5'-TTTGGGAGTTTTCACCAAGG-3'). One primer was both antisense and sense (as) because it is located in the inverted repeat of the *P* element. It is (5'-TGATGAATAACATAAGGTGGTCCGTCG-3'). RT primers are shown on the transgene map (Supplementary Fig. 9). qPCR was then performed on triplicates of each RT with a primer pair specific for the *nanos* gene in order to quantify the *nanos* transcripts. Simultaneously, qPCR was performed on triplicates of the same RT using different primer pairs corresponding to the former five regions of interest of *P{lacW}*. qPCR primer sequences are: 5'*P* (5'-CTGCAAAGCTGTGACTGGAG-3' and 5'-TTTGGGAGTTTTCACCAAGG-3'), 5'*lacZ* (5'-GAGAATCCGACGGTGTGTA-3' and 5'-AAATTCAGACGGCAAACGAC-3'), 3'*lacZ* (5'-ACTATCCCGACGCGCTTACT-3' and 5'-GTGGGCCATAATTCAATTCG-3'), 5'*white* (5'-GTCAATGTCCGCTTCACTT-3' and 5'-GGAGTTTGGCACAGCACTT-3') and 3'*P* (5'-CCACGGACATGCTAAGGGTTAA-3' and 5'-GTCGGCAAGAGACATCCACT-3'). The same series of dilutions composed of a mix of different RT preparations was used to normalize the quantity of *nanos* transcripts in all RT preparations leading to standard quantity (Sq) values for *nanos* transcripts in specific RT (using *nanos* primer = Sq(*nanos*)) or in control RT (without primer = Sq(control *nanos*)) preparations. A series of dilutions of a plasmid containing the *P{lacW}* transgene was used to normalize the quantity of transcripts of the clusters leading to Sq values for cluster transcript (Sq(specific) and Sq(control specific)). Variations between technical triplicates seem to be very low when compared to variations between biological replicates. The mean of the three technical replicates was then systematically used ($\bar{S}q$). The measure of the quantity of transcripts from a given region for one biological sample was then calculated using the formula: $(\bar{S}q(\text{specific}) - \bar{S}q(\text{control specific})) / (\bar{S}q(\text{nanos}) - \bar{S}q(\text{control nanos}))$. This allowed us to eliminate the background noise due to both sense and antisense transcripts (Sq(control transcript)) and to take into account variations in the quantity of RNA between biological samples (Sq(*nanos*)).

31. O'Kane, C. J. & Gehring, W. J. Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc. Natl Acad. Sci. USA* **84**, 9123–9127 (1987).

32. Bier, E. *et al.* Searching for pattern and mutation in the *Drosophila* genome with a P-lacZ vector. *Genes Dev.* **3**, 1273–1287 (1989).
33. Schupbach, T. & Wieschaus, E. Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129**, 1119–1136 (1991).
34. Wilson, J. E., Connell, J. E., Schlenker, J. D. & Macdonald, P. M. Novel genetic screen for genes involved in posterior body patterning in *Drosophila*. *Dev. Genet.* **19**, 199–209 (1996).
35. Harris, A. N. & Macdonald, P. M. *aubergine* encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. *Development* **128**, 2823–2832 (2001).
36. Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
37. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics

Roland Schmitz^{1*}, Ryan M. Young^{1*}, Michele Ceribelli^{1*}, Sameer Jhavar^{1*}, Wenming Xiao^{2*}, Meili Zhang¹, George Wright³, Arthur L. Shaffer¹, Daniel J. Hodson¹, Eric Buras¹, Xuelu Liu², John Powell², Yandan Yang¹, Weihong Xu¹, Hong Zhao¹, Holger Kohlhammer¹, Andreas Rosenwald⁴, Philip Kluij⁵, Hans Konrad Müller-Hermelink⁴, German Ott⁶, Randy D. Gascoyne⁷, Joseph M. Connors⁷, Lisa M. Rimsza⁸, Elias Campo⁹, Elaine S. Jaffe¹⁰, Jan Delabie¹¹, Erlend B. Smeland¹², Martin D. O'Gwang¹³, Steven J. Reynolds¹⁴, Richard I. Fisher¹⁵, Rita M. Brazier¹⁶, Raymond R. Tubbs¹⁷, James R. Cook¹⁷, Dennis D. Weisenburger¹⁸, Wing C. Chan¹⁸, Stefania Pittaluga¹⁰, Wyndham Wilson¹, Thomas A. Waldmann¹, Martin Rowe¹⁹, Sam M. Mbulaiteye²⁰, Alan B. Rickinson¹⁹ & Louis M. Staudt¹

Burkitt's lymphoma (BL) can often be cured by intensive chemotherapy, but the toxicity of such therapy precludes its use in the elderly and in patients with endemic BL in developing countries, necessitating new strategies¹. The normal germinal centre B cell is the presumed cell of origin for both BL and diffuse large B-cell lymphoma (DLBCL), yet gene expression analysis suggests that these malignancies may use different oncogenic pathways². BL is subdivided into a sporadic subtype that is diagnosed in developed countries, the Epstein-Barr-virus-associated endemic subtype, and an HIV-associated subtype, but it is unclear whether these subtypes use similar or divergent oncogenic mechanisms. Here we used high-throughput RNA sequencing and RNA interference screening to discover essential regulatory pathways in BL that cooperate with *MYC*, the defining oncogene of this cancer. In 70% of sporadic BL cases, mutations affecting the transcription factor *TCF3* (E2A) or its negative regulator *ID3* fostered *TCF3* dependency. *TCF3* activated the pro-survival phosphatidylinositol-3-OH kinase pathway in BL, in part by augmenting tonic B-cell receptor signalling. In 38% of sporadic BL cases, oncogenic *CCND3* mutations produced highly stable cyclin D3 isoforms that drive cell cycle progression. These findings suggest opportunities to improve therapy for patients with BL.

We performed RNA resequencing (RNA-seq) on 28 sporadic BL patient biopsies and 13 BL cell lines and reanalysed published RNA-seq data from 52 germinal centre B-cell-like (GCB) DLBCL cases and 28 activated B-cell-like (ABC) DLBCL cases³. Elimination of known single-nucleotide polymorphisms left a set of putative single-nucleotide variants, of which 95% (495 out of 518) were confirmed by Sanger sequencing (Supplementary Tables 1 and 2).

Mutations in many genes were more frequent in BL than in DLBCL, including *MYC*, as well as many not previously implicated in this lymphoma subtype (Fig. 1, Supplementary Fig. 1a and Supplementary Table 1). Conversely, recurrently mutated genes in DLBCL³⁻⁷ (*EZH2*, *SGK1*, *BCL2*, *CD79B*, *MYD88*) were rarely, if ever, mutated in BL. Several genes were mutated in BL and DLBCL (*TP53*, *GNA13*, *MKI67*, *CCND3*), although *TP53* mutations were more common in BL

(Fig. 1 and Supplementary Fig. 1b). This mutational survey indicates that BL is pathogenetically distinct from other germinal centre-derived lymphomas.

Highly recurrent mutations in *TCF3* and its negative regulator *ID3* indicated that *TCF3* has a central role in BL pathogenesis, as it does in normal B-cell development by regulating the transcription of immunoglobulin and other B-cell-restricted genes through E-box motifs^{8,9}. *ID3* and/or *TCF3* mutations were present in sporadic BL, HIV-associated BL and endemic BL, in 70%, 67% and 40% of samples, respectively, but these mutations were rare in other lymphoid cancers (Fig. 2a and Supplementary Table 3). In sporadic BL, *ID3* mutations (58%) were more common than *TCF3* mutations (11%), and some tumours had mutations in both genes (13%). *ID3* mutations were usually bi-allelic whereas *TCF3* mutations were often mono-allelic (Supplementary Fig. 2a and Supplementary Table 3). A somatic origin was confirmed for 14 *ID3* mutations and 4 *TCF3* mutations

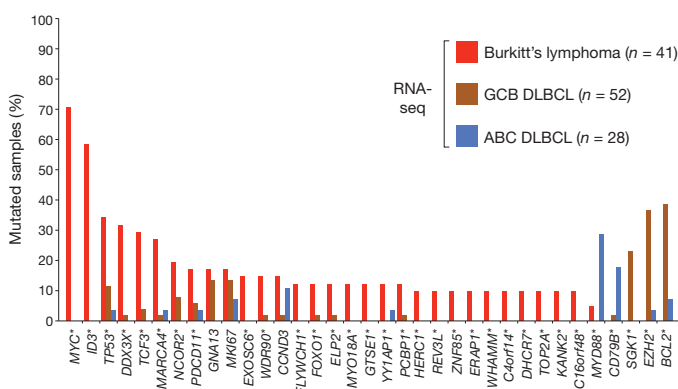


Figure 1 | Recurrently mutated genes in aggressive lymphomas determined by RNA-seq. Shown are genes that were recurrently mutated in BL based on RNA-seq analysis (≥ 4 out of 41 samples), as well as representative genes known to be recurrently mutated in DLBCL. Asterisks indicate differentially mutated genes ($P < 0.05$; Supplementary Table 9).

¹Metabolism Branch Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA. ²Bioinformatics and Molecular Analysis Section, Division of Computational Bioscience, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892, USA. ³Biometric Research Branch, DCTD, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA. ⁴Department of Pathology, University of Würzburg, 97080 Würzburg, Germany. ⁵Department of Pathology and Medical Biology, Groningen University Medical Center, University of Groningen, 9713 GZ Groningen, The Netherlands. ⁶Department of Clinical Pathology, Robert-Bosch-Krankenhaus, and Dr. Margarete Fischer-Bosch Institute for Clinical Pharmacology, 70376 Stuttgart, Germany. ⁷British Columbia Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada. ⁸Department of Pathology, University of Arizona, Tucson, Arizona 85724, USA. ⁹Hospital Clinic, University of Barcelona, 08036 Barcelona, Spain. ¹⁰Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA. ¹¹Pathology Clinic, Rikshospitalet University Hospital, 0372 Oslo, Norway. ¹²Institute for Cancer Research, Rikshospitalet University Hospital and Center for Cancer Biomedicine, Faculty Division of the Norwegian Radium Hospital, University of Oslo, 0310 Oslo, Norway. ¹³St. Mary's Hospital Lacor, Gulu 256, Uganda. ¹⁴Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹⁵James P. Wilmot Cancer Center, University of Rochester School of Medicine, Rochester 14642, New York, USA. ¹⁶Oregon Health and Science University, Portland, Oregon 97239, USA. ¹⁷Cleveland Clinic Pathology and Laboratory Medicine Institute, Cleveland, Ohio 44195, USA. ¹⁸Departments of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA. ¹⁹School of Cancer Sciences, Birmingham Cancer Research UK Centre, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. ²⁰Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Department of Health and Human Services, Rockville, Maryland 20852, USA.

*These authors contributed equally to this work.

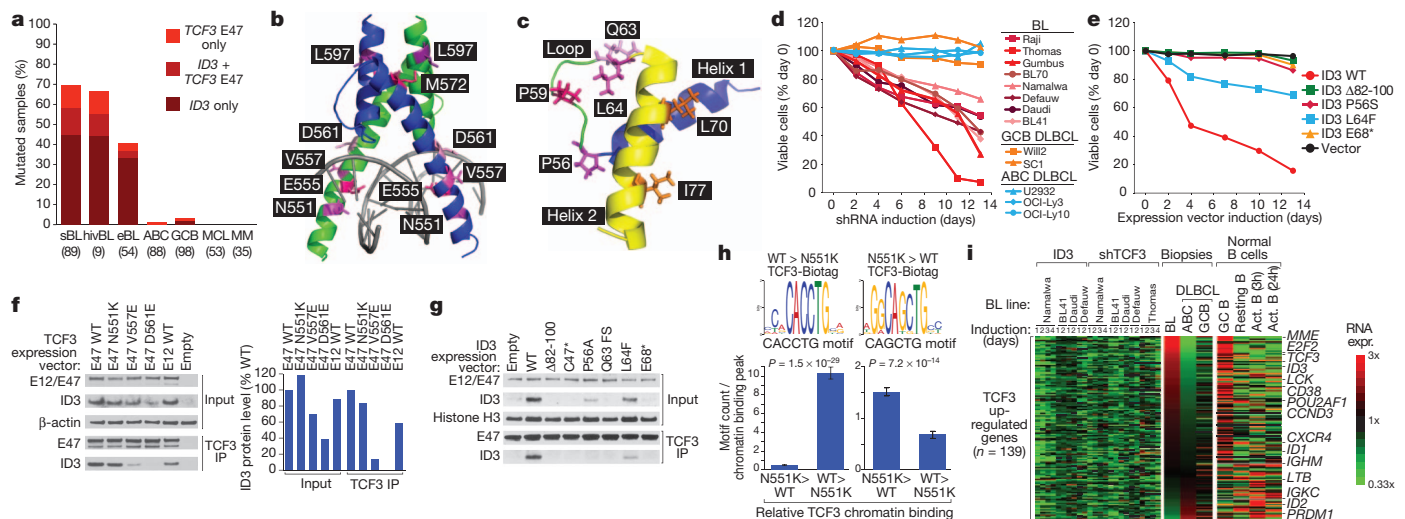


Figure 2 | *TCF3* is essential for Burkitt lymphoma viability. **a**, *TCF3* and *ID3* mutation frequencies in lymphoid cancers. eBL, endemic BL; hivBL, HIV-associated BL; sBL, sporadic BL; MCL, mantle cell lymphoma; MM, multiple myeloma. **b**, Location of BL mutants in the crystal structure of the dimeric *TCF3* E47 B-HLH domain¹⁰. **c**, Location of BL mutants in the crystal structure of the *ID3* HLH domain (Protein Data Bank accession number 2LFH). **d**, Selective toxicity of a *TCF3* shRNA for BL lines. Shown is the fraction of GFP⁺, shRNA-expressing cells relative to the GFP⁺, shRNA-negative fraction at the indicated times, normalized to the day 0 values. Data are representative of four experiments. **e**, Toxicity of wild-type (WT) but not mutant *ID3* isoforms for the *ID3*-mutant Namalwa BL line. Shown is the fraction of GFP⁺, *ID3*-expressing cells relative to the GFP⁺, *ID3*-negative cells, normalized to the day 0 values. Data are representative of four experiments. **f**, *TCF3* mutants with reduced ability to bind *ID3*. WT or mutant *TCF3* isoforms were coexpressed with WT *ID3* in 293T cells. The indicated proteins were detected in total cellular extracts (input) or after anti-*TCF3* immunoprecipitation (IP) (left). **g**, *ID3* levels were quantified by densitometry and normalized to *TCF3* E47 levels (right). **h**, BL-derived mutant *ID3* proteins are less stable than WT *ID3* and bind *TCF3* less well. Mutant or WT *ID3* isoforms were expressed in the *ID3*-deficient Namalwa BL line. The indicated proteins were detected in total cellular extracts (input) or after anti-*TCF3* IP. **i**, *TCF3*(N551K) is an altered-specificity mutant. Shown on top are DNA base frequencies of the most enriched motifs in peaks bound more than fourfold more or less by *TCF3*(N551K) compared to WT *TCF3*. The mean number (\pm s.e.m.) of the indicated motifs per differentially bound peak is plotted below. **j**, A *TCF3* gene expression signature expressed in BL and normal germinal centre B cells. Gene expression changes were profiled in *ID3*-mutant BL lines following *TCF3* knockdown or WT *ID3* overexpression. Shown are genes that were downregulated by at least 0.33 log₂ in >70% of samples. Average expression of these genes in the indicated lymphoma subtypes based on published data² and in B-cell subpopulations based on RNA-seq is shown.

(Supplementary Table 3). All *TCF3* mutations affected the basic helix–loop–helix (B-HLH) DNA-binding and dimerization domain of one *TCF3* splice isoform (E47) but not the other (E12), indicating a non-redundant role for E47 in BL pathogenesis. In cases with *TCF3* mutations, E47 was more highly expressed than E12, indicating gain-of-function (Supplementary Fig. 2b).

Most *TCF3* mutations target four evolutionarily conserved residues in the B-HLH region (N551K, V557E/G, D561E/V/N, M572K; Supplementary Fig. 3a). The most common mutations affect V557 and D561, which are adjacent in the crystal structure and face away from DNA, indicating a role in intermolecular interactions (Fig. 2b). The B-HLH domain may be distorted by mutations affecting M572 and L597, which are neighbouring residues in the crystal structure. N551 is a DNA contact residue¹⁰, indicating that N551K could alter *TCF3* DNA binding.

A variety of nonsense and frameshift mutations inactivate *ID3* in BL tumours, suggesting a tumour-suppressor mechanism (Supplementary Fig. 3b). Many missense mutations target the conserved loop region of *ID3*, potentially changing the tertiary structure of the B-HLH domain and impairing its ability to inhibit *TCF3* (ref. 11, Fig. 2c). Numerous *ID3* missense mutations affect the HLH domain away from the interface of the two helices, possibly altering *TCF3* interaction. Other mutations disrupt an *ID3* splice donor and force a cryptic splice donor to be used, thereby deleting residues V82–Q100 (Supplementary Figs 2c and 3b).

An RNA interference screen revealed *TCF3* to be an essential gene in BL lines (Supplementary Fig. 2d and Supplementary Table 4), supporting the notion that the *TCF3* and *ID3* mutations in BL promote *TCF3* action. *TCF3* knockdown caused a time-dependent toxicity in all BL lines, irrespective of *ID3*/*TCF3* mutations, but had no effect on DLBCL lines (Fig. 2d and Supplementary Figs 2e and 4a, b). Wild-type *TCF3* rescued BL lines from the toxicity of RNA-interference-mediated

depletion of *TCF3*, as did the *TCF3* mutants, indicating that they are not loss-of-function (Supplementary Fig. 2f). Introduction of wild-type *ID3* into BL lines with *ID3* mutations was lethal, but BL-derived *ID3* mutants had less or no toxicity, consistent with a tumour-suppressor mechanism (Fig. 2e and Supplementary Figs 2g and 4c). The common *ID3* and *TCF3* mutants diminished their inhibitory heterodimerization. *TCF3*(V557E) and *TCF3*(D561E) did not associate well with *ID3* and failed to stabilize *ID3* protein expression, unlike wild-type *TCF3* (Fig. 2f and Supplementary Figs 2h and 4d). Likewise, the *ID3* mutant proteins were expressed less well than wild-type *ID3* and were less able to co-immunoprecipitate *TCF3* (Fig. 2g and Supplementary Fig. 4e). However, *TCF3*(N551K) behaved like wild-type *TCF3* in these dimerization assays, indicating a distinct mechanism.

We next used chromatin immunoprecipitation followed by sequencing (ChIP-seq) analysis to gauge the ability of the *TCF3* mutants to interact with chromatin genome-wide. We engineered two BL lines to express biotinylated wild-type or mutant *TCF3* isoforms (*TCF3*-Biotag; see Methods), allowing us to precipitate bound chromatin with streptavidin. For comparison, we used anti-*TCF3* antibodies to precipitate chromatin in unmanipulated BL cells (Supplementary Table 5). Both the endogenous and *TCF3*-Biotag ChIP-seq peaks were enriched for E-box motifs (CAG(G/C)TG) and overlapped extensively (Supplementary Fig. 5a). In 25-base-pair bins bound by wild-type *TCF3*-Biotag, the V557E, D561E and N551K isoforms had overlapping ChIP-seq tags (>7) in 98%, 98% and 92% of instances, respectively, but the overlap was only 10% for control Biotag ChIP-seq data ($P < 10^{-300}$). Hence, all *TCF3* mutants bound a large number of genomic targets equivalently.

Given the lower overlap between *TCF3*(N551K) and wild-type *TCF3* chromatin binding, we identified genomic regions that had

fourfold greater ($n = 212$) or lesser ($n = 139$) association with wild-type TCF3 than TCF3(N551K) ($P < 10^{-10}$) (Supplementary Table 6 and Supplementary Fig. 5b). In these binding regions, TCF3(V557E) and TCF3(D561E) behaved like wild-type TCF3. The peaks bound preferentially by wild-type TCF3 contained multiple copies of the motif 5'-NNCACCTG-3' whereas the peaks bound preferentially by TCF3(N551K) were enriched for the sequence 5'-GGCAGCTG-3' (Fig. 2h). Whereas both motifs match the E-box consensus, these results indicate that TCF3(N551K) is an altered-specificity mutant that has somewhat different genomic targets than wild-type TCF3.

To gain insight into the biological processes controlled by TCF3 in BL, we profiled changes in gene expression following TCF3 knock-down and following wild-type ID3 expression in ID3-mutant BL lines. We identified 139 'TCF3-upregulated' genes that were decreased in expression by both manipulations and 166 'TCF3-downregulated' genes that were increased in expression (false discovery rate (FDR) = 0.017; Fig. 2i and Supplementary Figs 2i and 6a). TCF3 ChIP-seq peaks were enriched among TCF3-upregulated genes (58%; $P = 1.81 \times 10^{-29}$) and among TCF3-downregulated genes (32%; $P = 1.03 \times 10^{-4}$) (Supplementary Fig. 6a). We will refer to such genes as 'TCF3 direct targets'.

Most TCF3-upregulated genes were more highly expressed in BL than in DLBCL, whereas TCF3-downregulated genes were generally expressed at lower levels in BL ($P \leq 0.001$; Fig. 2i and Supplementary Fig. 6a, b). BL tumours with ID3 and/or TCF3 mutations had higher expression of the TCF3-upregulated signature than tumours with wild-type ID3 and TCF3, and the opposite was true for the TCF3-downregulated signature ($P = 0.0001$; Supplementary Fig. 6c). Hence, the transcriptional influence of TCF3 on the BL phenotype seems to be accentuated by ID3/TCF3 mutations. TCF3-upregulated genes were more highly expressed in germinal centre B cells than in resting or activated blood B cells, and the reverse was true for TCF3-downregulated genes (Fig. 2i and Supplementary Fig. 6a), indicating that BL 'inherits' the TCF3 gene expression program from its normal cellular counterpart.

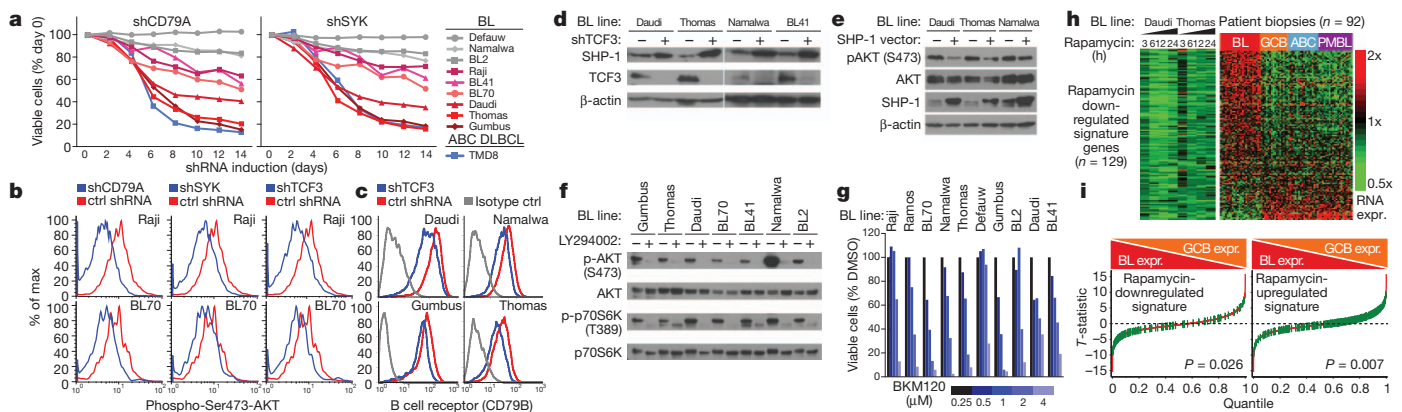


Figure 3 | Tonic BCR signalling and PI(3) kinase activity in Burkitt's lymphoma. **a**, CD79A and SYK shRNAs are toxic for a subset of BL lines. Shown is the fraction of GFP⁺, shRNA-expressing cells relative to the GFP⁺, shRNA-negative fraction at the indicated times, normalized to the day 0 values. BCR-dependent BL lines are depicted using red colours. The BCR-dependent ABC DLBCL line TMD8 (ref. 4) is also shown. **b**, Knockdown of CD79A, SYK or TCF3 reduces PI(3) kinase activity. Following induction of the indicated shRNAs for 2 days, shRNA-expressing (GFP⁺) cells were analysed by fluorescence-activated cell sorting (FACS) for phospho-S473-AKT as a measure of PI(3) kinase activity. **c**, TCF3 regulates surface BCR expression in BL. Following induction of the indicated shRNAs for 1 day, surface BCR expression (CD79B) was quantified by FACS in shRNA-expressing (GFP⁺) cells. **d**, TCF3 suppresses PTPN6 (SHP-1) expression. A TCF3 shRNA was induced in BL lines for 2 days, followed by immunoblotting for the indicated proteins. **e**, SHP-1 suppresses phospho-S473-AKT in BL lines. BL lines were transduced with a SHP-1 expression vector

Biological insights from this analysis include the fact that the negative regulators of TCF3—ID1, ID2 and ID3—were direct targets of TCF3 transactivation, thereby creating a negative feedback loop (Fig. 2i and Supplementary Fig. 5a). By RNA-seq, ID3 was 38-fold and 12-fold more highly expressed in BL than ID1 and ID2, respectively, accounting for the preferential mutation of ID3 in BL. TCF3 also positively regulated genes that have crucial roles in germinal centre B-cell biology (POU2AF1, CXCR4, LTB, CCND3). TCF3 upregulated CCND3 and E2F2 while downregulating RB1, thereby promoting cell cycle progression.

Two components of the B-cell receptor (BCR), the immunoglobulin heavy and light chains, were both upregulated by TCF3 in BL, as in normal B cells^{8,12} (Fig. 2i and Supplementary Figs 2i, j and 5a). In this regard, it was notable that knockdown of the BCR subunit CD79A was toxic for several BL lines in our RNA interference screen (Supplementary Fig. 7a). Two-thirds of BL lines were clearly BCR-dependent, on the basis of a time-dependent decrease in their viability following knockdown of either CD79A or the BCR-associated kinase SYK (Fig. 3a). Unlike ABC DLBCL lines, which have a 'chronic active' form of BCR signalling⁴, BL lines do not require the NF- κ B pathway for survival because they were not killed by an I κ B kinase β inhibitor and had little or no dependence on CARD11 (ref. 13), an adaptor that engages NF- κ B (Supplementary Fig. 7a–c). Rather, CD79A or SYK depletion in BL lines decreased AKT phosphorylation, a marker of phosphatidylinositol-3-OH (PI(3)) kinase signalling (Fig. 3b and Supplementary Fig. 7d), indicating that the BCR-dependency in BL is akin to 'tonic' BCR signalling¹⁴, a phenomenon that engages pro-survival PI(3) kinase signalling more than NF- κ B¹⁵.

TCF3 knockdown decreased phospho-AKT levels in all BCR-dependent lines tested, as did ID3 overexpression (Fig. 3b and Supplementary Fig. 7d, e), perhaps due to decreased cell-surface BCR expression following TCF3 depletion (Fig. 3c). In addition, a direct TCF3 target, PTPN6, encodes the phosphatase SHP-1, an inhibitor of BCR signalling (Supplementary Figs 2j and 5a). TCF3 depletion increased SHP-1 mRNA and protein levels, indicating

(+) or empty vector (–), whereupon the indicated proteins were analysed by immunoblotting. **f**, BL lines have constitutively PI(3) kinase activity. The indicated proteins were analysed by immunoblotting, before and after treatment with the PI(3) kinase inhibitor LY294002. **g**, PI(3) kinase inhibition is toxic to BL lines. Viable BL cells were quantified by MTS assay following treatment for 4 days with the indicated concentrations of the pan-class I PI(3) kinase inhibitor BKM120. **h**, A signature of rapamycin-responsive genes is highly expressed in BL. Changes of gene expression were profiled over time in 2 BL lines following rapamycin (100 pM) treatment. Genes consistently downregulated in both lines were chosen (see Methods), and their expression in lymphoma biopsies² is shown based on the colour scale. PMBL, primary mediastinal B-cell lymphoma. **i**, The rapamycin-upregulated and -downregulated signatures distinguish BL and GCB DLBCL. Genes are ranked according to their expression in BL versus GCB DLBCL (T -statistic) and rapamycin signature genes are indicated with a green hash mark. Kolmogorov–Smirnov P values are shown.

TCF3 repression (Fig. 3d and Supplementary Figs 2i and 6a). Ectopic provision of SHP-1 decreased phospho-AKT in BCR-dependent BL lines, indicating that TCF3 repression of SHP-1 may contribute to tonic BCR signalling and PI(3) kinase activation in BL (Fig. 3e).

A screen of a larger number of BL lines revealed that all had PI(3) kinase-dependent AKT phosphorylation and engagement of the mTOR pathway, as judged by phosphorylation of p70 S6 kinase (Fig. 3f). Treatment of BL lines with BKM120, a PI(3) kinase inhibitor in clinical trials, or rapamycin, an inhibitor of the mTORC1 complex, was toxic to most BL lines (Fig. 3g and Supplementary Fig. 7f). Of note, both BCR-dependent and -independent lines had constitutive PI(3) kinase signalling. Other mechanisms to activate PI(3) kinase in BL include *PTEN* mutations, which were infrequent (7%), and tenfold overexpression (compared to DLBCL) of the MYC-dependent gene

MIR17HG, which encodes a microRNA that inhibits *PTEN* expression¹⁶ (Supplementary Fig. 7g). To judge whether the PI(3) kinase pathway may be active in primary BL tumours, we identified genes that were significantly up- or downregulated following rapamycin treatment of BL lines (FDR = 0.0022; Supplementary Table 7). Among BL and GCB DLBCL biopsies, rapamycin-downregulated genes were generally more highly expressed in BL ($P = 0.026$), whereas rapamycin-upregulated genes had the opposite enrichment pattern ($P = 0.007$) (Fig. 3i), indicating that PI(3) kinase-dependent mTORC1 activity is a consistent feature of BL tumours.

Another aspect of BL pathogenesis was revealed by recurrent mutations in the TCF3 direct target *CCND3*, encoding cyclin D3, a required regulator of the G1–S cell cycle transition in germinal centre B cells^{17,18}. *CCND3* mutations were frequent in sporadic BL (38%) and HIV-associated BL (67%) but not endemic BL (1.8%), indicating a distinct genetic pathogenesis for this latter BL subtype (Fig. 4b). At a lower frequency, *CCND3* mutations were also present in ABC and GCB DLBCL^{3,6,7}. Multiple nonsense and frameshift mutations removed up to 41 amino acids from the cyclin D3 C terminus (Fig. 4a and Supplementary Table 8). Recurrent missense mutations affected threonine 283 (T283), known to be involved in D-type cyclin phosphorylation and stability¹⁹, as well as nearby proline (P284) and isoleucine (I290) residues. These cyclin D3 residues were conserved in evolution, and similar residues are present in cyclin D1 and D2 (Fig. 4a). Most mutations were heterozygous and their somatic origin was confirmed in five cases (Supplementary Table 8).

To explore the function of the cyclin D3 mutants, we constructed fusion proteins linking green fluorescent protein (GFP) to either wild-type or mutant cyclin D3. All mutant isoforms accumulated to more than tenfold higher levels than the wild-type isoform (Fig. 4c), and pulse-chase analysis showed that the mutant cyclin D3 isoforms have longer half lives (Supplementary Fig. 8a). To test the oncogenic potential of the cyclin D3 mutants, we transduced GFP-tagged wild-type or T283A cyclin D3 into lymphoma lines in which endogenous cyclin D3 was knocked down. Cells transduced with T283A cyclin D3 had a marked proliferative advantage over untransduced cells, but wild-type cyclin D3 had little effect (Fig. 4d). Separately, our RNA interference screen revealed that BL and GCB DLBCL lines depend on cyclin D3 and CDK6, a kinase that partners with D-type cyclins, irrespective of

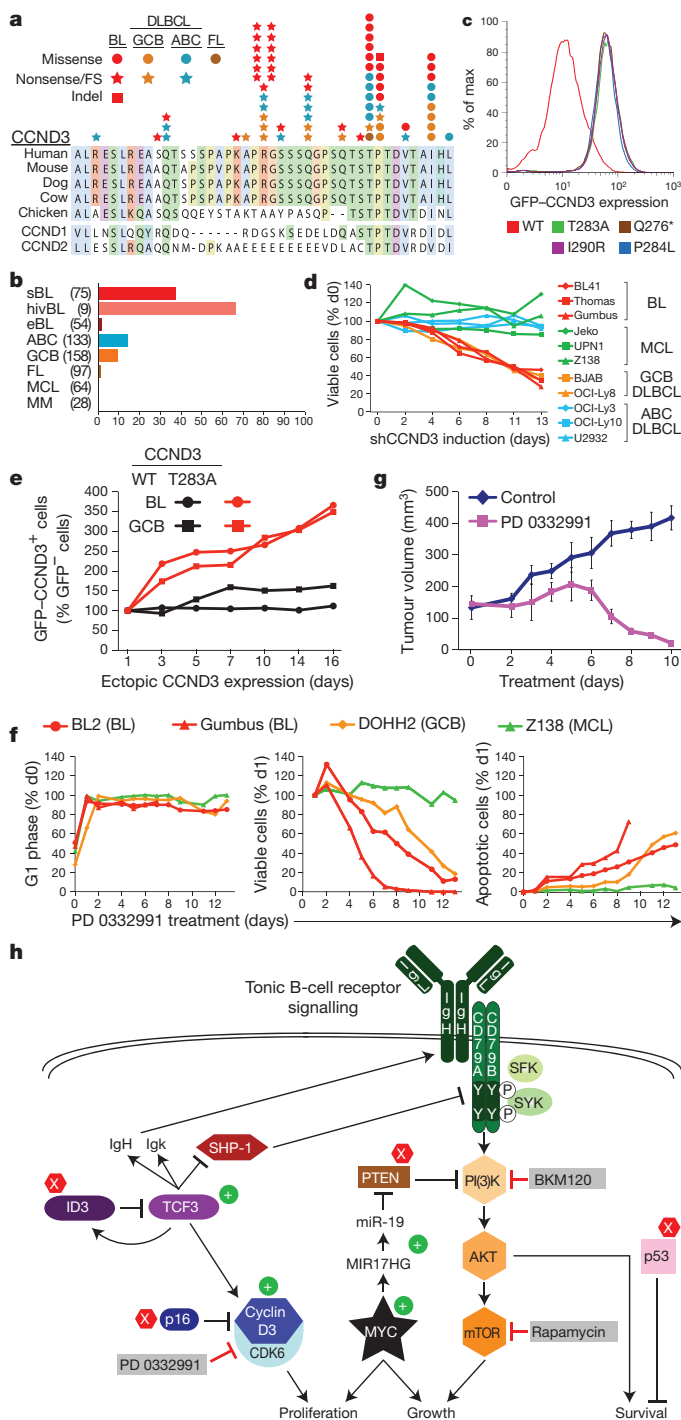


Figure 4 | Oncogenic *CCND3* mutations in Burkitt's lymphoma. **a**, Cyclin D3 residues affected by the indicated mutations in each lymphoma subtype. Amino acids 250–292 of NP_001751 are shown. FL, follicular lymphoma; FS, frameshift. **b**, Frequencies of *CCND3* mutations in different lymphoma subtypes. **c**, *CCND3* mutations increase protein stability. FACS analysis of the Gumbus BL line transduced with WT or mutant GFP–*CCND3* fusion proteins. **d**, The T283A cyclin D3 mutant confers a proliferation advantage. Expression of endogenous *CCND3* was knocked down in Gumbus (BL) and BJAB (GCB DLBCL) cells and different GFP–*CCND3* isoforms were ectopically expressed. The relative number of GFP–*CCND3*-expressing cells is plotted over time of shRNA and GFP–*CCND3* induction, normalized to day 0. Data are representative of three experiments. **e**, *CCND3* shRNAs are selectively toxic for BL and GCB DLBCL lines. Shown is the fraction of GFP⁺, shRNA-expressing cells relative to the GFP⁺, shRNA-negative fraction at the indicated times, normalized to the day 0 values. Data are representative of four experiments. **f**, Cell cycle block in G1 phase is lethal to cyclin D3-mutant lymphoma lines. Lines were treated with the CDK4/6 inhibitor PD 0332991 (1 μ M) over the indicated time course and analysed for viable cells in G1 phase, total viable cells and apoptotic cells. Data are normalized as indicated and are representative of three experiments. **g**, Therapeutic potential of PD 0332991 revealed using a BL xenograft model. Immunodeficient mice bearing established subcutaneous xenografts of the Gumbus BL line (engineered to express luciferase) were treated with PD 0332991 (150 mg per kg per day per os) for the indicated times. Tumour volumes were estimated by luciferase luminescence. Error bars are s.e.m. ($n = 3$). **h**, Schematic of recurrent oncogenic pathways in Burkitt's lymphoma. Gain-of-function and loss-of-function aberrations are indicated by + signs and by \times signs, respectively. Grey boxes indicate drugs that block these deregulated pathways.

CCND3 mutational status (Fig. 4e, Supplementary Fig. 8b–d and Supplementary Table 4). Hence, BL lines rely on cyclin D3/CDK6 for cell cycle progression, an effect augmented by oncogenic cyclin D3 mutations. The BL cell cycle is also deregulated by nonsense and frameshift mutations or homozygous deletions in *CDKN2A*, encoding the CDK6 inhibitor p16 (Supplementary Fig. 8e and Supplementary Table 8).

To explore this pathway as a therapeutic target, we treated BL, GCB DLBCL and mantle cell lymphoma (MCL) lines with a CDK4/6 inhibitor (PD 0332991) daily for 2 weeks. After an arrest in G1 phase, the BL and GCB DLBCL lines began to die by day 2, with a steady accumulation of apoptotic cells over time, whereas the MCL line arrested in G1 phase but did not die (Fig. 4f). Treatment of a BL xenograft model after the establishment of tumours with PD 0332991 profoundly reduced tumour volume after 6 days, resulting in the virtual disappearance of tumour cells by day 10 (Fig. 4g and Supplementary Fig. 8f).

By merging functional and structural genomic data we have uncovered previously unappreciated pathways in BL pathogenesis, several of which are amenable to therapeutic attack (Fig. 4h). The majority of BL tumours acquire mutations that free TCF3 from ID3 inhibition. These mutations 'hard-wire' a TCF3 transcriptional program that is characteristic of germinal centre B cells and distinguishes BL from other aggressive lymphomas. BL lines require TCF3 for survival, in part because it augments pro-survival PI(3) kinase signalling by intensifying a tonic form of BCR signalling. The oncogenic synergy between the MYC and PI(3) kinase pathways that is suggested by our study is supported by the generation of BL-like tumours in mice in which these two pathways are deregulated²⁰. Additionally, the key role of cyclin D3/CDK6 in BL pathogenesis is reinforced by the identification of cyclin D3 mutants in this mouse model.

Whereas high-dose chemotherapy can often cure BL in younger patients from developed countries¹, these regimens are unsafe in older patients and cannot be deployed in less developed regions due to immune suppression and to logistical difficulties that preclude effective delivery²¹. Hopefully, the new insights into BL pathogenesis described herein will prompt clinical evaluation of drugs targeting the PI(3) kinase pathway, tonic BCR signalling, and cyclin D3/CDK6 in BL. Eventually, the rational combination of such targeted agents could provide more effective and less-toxic treatment of BL worldwide.

METHODS SUMMARY

RNA-Seq was performed using established Illumina protocols on a HiSeq 2000 sequencer. RNA interference screening and cellular toxicity assays were conducted as described^{5,13}. Gene expression profiling was performed using Agilent 4 × 44K microarrays. Detailed experimental and analytic procedures are presented in Supplementary Methods.

Received 25 January; accepted 11 July 2012.

Published online 12 August 2012.

1. Yustein, J. T. & Dang, C. V. Biology and treatment of Burkitt's lymphoma. *Curr. Opin. Hematol.* **14**, 375–381 (2007).
2. Dave, S. S. *et al.* Molecular diagnosis of Burkitt's lymphoma. *N. Engl. J. Med.* **354**, 2431–2442 (2006).
3. Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298–303 (2011).
4. Davis, R. E. *et al.* Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* **463**, 88–92 (2010).
5. Ngo, V. N. *et al.* Oncogenically active *MYD88* mutations in human lymphoma. *Nature* **470**, 115–119 (2011).

6. Pasqualucci, L. *et al.* Analysis of the coding genome of diffuse large B-cell lymphoma. *Nature Genet.* **43**, 830–837 (2011).
7. Lohr, J. G. *et al.* Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc. Natl Acad. Sci. USA* **109**, 3879–3884 (2012).
8. Murre, C., McCaw, P. S. & Baltimore, D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD*, and *myc* proteins. *Cell* **56**, 777–783 (1989).
9. Kee, B. L. E and ID proteins branch out. *Nature Rev. Immunol.* **9**, 175–184 (2009).
10. Ellenberger, T., Fass, D., Arnaud, M. & Harrison, S. C. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev.* **8**, 970–980 (1994).
11. Pesce, S. & Benezra, R. The loop region of the helix-loop-helix protein Id1 is critical for its dominant negative activity. *Mol. Cell. Biol.* **13**, 7874–7880 (1993).
12. Murre, C. *et al.* Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**, 537–544 (1989).
13. Ngo, V. N. *et al.* A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441**, 106–110 (2006).
14. Lam, K. P., Kuhn, R. & Rajewsky, K. *In vivo* ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death. *Cell* **90**, 1073–1083 (1997).
15. Srinivasan, L. *et al.* PI3 kinase signals BCR-dependent mature B cell survival. *Cell* **139**, 573–586 (2009).
16. Xiao, C. *et al.* Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. *Nature Immunol.* **9**, 405–414 (2008).
17. Cato, M. H., Chintalapati, S. K., Yau, I. W., Omori, S. A. & Rickert, R. C. Cyclin D3 is selectively required for proliferative expansion of germinal center B cells. *Mol. Cell. Biol.* **31**, 127–137 (2010).
18. Peled, J. U. *et al.* Requirement for cyclin D3 in germinal center formation and function. *Cell Res.* **20**, 631–646 (2010).
19. Diehl, J. A., Cheng, M., Roussel, M. F. & Sherr, C. J. Glycogen synthase kinase-3 β regulates cyclin D1 proteolysis and subcellular localization. *Genes Dev.* **12**, 3499–3511 (1998).
20. Sander, S. *et al.* Synergy between PI3K signalling and MYC in Burkitt lymphomagenesis. *Cancer Cell* <http://dx.doi.org/10.1016/j.ccr.2012.06.012> (2012).
21. Orem, J., Mbidde, E. K. & Weiderpass, E. Current investigations and treatment of Burkitt's lymphoma in Africa. *Trop. Doct.* **38**, 7–11 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, an NCI SPECS grant (U01-CA 114778), by the Foundation for NIH, through a gift from the Richard A. Lauderbaugh Memorial Fund, and by Cancer Research UK. This study was conducted under the auspices of the Lymphoma/Leukemia Molecular Profiling Project (LLMPP). R.S. was supported by the Dr Mildred Scheel Stiftung für Krebsforschung (Deutsche Krebshilfe). D.J.H. is a Kay Kendall Leukaemia Fund Intermediate research fellow. This study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health (<http://biowulf.nih.gov>). We thank K. Meyer for help with the GEO submission, T. Ellenberger for the TCF3 crystal structure coordinates, B. Tran (Center for Cancer Research Sequencing Facility) and K. Hartman for DNA sequencing and K. Rajewsky for discussions. The DLBCL data set is part of the Cancer Genomics Characterization Initiative (CGCI), supported by NCI contract N01-C0-12400 (<http://cgap.nci.nih.gov/cgci.html/>) and was obtained from dbGaP at <http://www.ncbi.nlm.nih.gov/gap>. We thank the participants in the EMBLEM Study (<http://emblem.cancer.gov/>) in Uganda, the EMBLEM Study staff for collecting and processing the samples and data, and the Government of Uganda for allowing the study to be done and samples to be exported for research.

Author Contributions R.S., R.M.Y., M.C., S.J., M.Z., H.K., A.L.S. and D.J.H. designed and performed experiments. T.A.W. designed experiments. W.Xu Y.Y., E.B. and H.Z. performed experiments. W.Xi., G.W., X.L. and J.P. analysed data. A.R., P.K., H.K.M.-H., G.O., R.D.G., J.M.C., L.M.R., E.C., E.S.J., J.D., E.B.S., R.I.F., R.M.B., R.R.T., J.R.C., D.D.W., W.C.C., S.P., W.W., M.D.O., S.J.R., S.M.M., M.R. and A.B.R. supplied BL patient samples or lines, and reviewed pathological and clinical data. L.M.S. designed and supervised research and wrote the manuscript.

Author Information Gene expression profiling data have been submitted to GEO under accession number GSE35163, RNA-seq data has been deposited in NCBI Sequence Read Archive (SRA048058) and ChIP-seq data has been deposited in NCBI Sequence Read Archive (SRA052618). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.M.S. (Istaudt@mail.nih.gov).

Structural basis for RNA–duplex recognition and unwinding by the DEAD–box helicase Mss116p

Anna L. Mallam¹, Mark Del Campo^{1†}, Benjamin Gilman¹, David J. Sidote¹ & Alan M. Lambowitz¹

DEAD-box proteins are the largest family of nucleic acid helicases, and are crucial to RNA metabolism throughout all domains of life^{1,2}. They contain a conserved ‘helicase core’ of two RecA-like domains (domains (D)1 and D2), which uses ATP to catalyse the unwinding of short RNA duplexes by non-processive, local strand separation³. This mode of action differs from that of translocating helicases and allows DEAD-box proteins to remodel large RNAs and RNA–protein complexes without globally disrupting RNA structure⁴. However, the structural basis for this distinctive mode of RNA unwinding remains unclear. Here, structural, biochemical and genetic analyses of the yeast DEAD-box protein Mss116p indicate that the helicase core domains have modular functions that enable a novel mechanism for RNA–duplex recognition and unwinding. By investigating D1 and D2 individually and together, we find that D1 acts as an ATP-binding domain and D2 functions as an RNA–duplex recognition domain. D2 contains a nucleic-acid-binding pocket that is formed by conserved DEAD-box protein sequence motifs and accommodates A-form but not B-form duplexes, providing a basis for RNA substrate specificity. Upon a conformational change in which the two core domains join to form a ‘closed state’ with an ATPase active site, conserved motifs in D1 promote the unwinding of duplex substrates bound to D2 by excluding one RNA strand and bending the other. Our results provide a comprehensive structural model for how DEAD-box proteins recognize and unwind RNA duplexes. This model explains key features of DEAD-box protein function and affords a new perspective on how the evolutionarily related cores of other RNA and DNA helicases diverged to use different mechanisms.

Mss116p is a DEAD-box RNA helicase that facilitates the folding and splicing of mitochondrial group I and group II introns primarily by acting as an RNA chaperone that unwinds RNA duplexes to disrupt stable but inactive RNA structures^{5–7}. The RecA-like helicase core domains of Mss116p, which together catalyse RNA unwinding^{8,9}, contain conserved DEAD-box protein sequence motifs that are required for helicase function (Fig. 1a). D2 also includes a non-conserved carboxy-terminal extension (CTE) that stabilizes the domain and extends its RNA-binding surface¹⁰. Small-angle X-ray scattering studies show that without substrates, the helicase core of Mss116p adopts an extended ‘open state’ conformation, as observed for other DEAD-box proteins^{11,12}. A compact ‘closed state’, the X-ray crystal structure of which has been determined for Mss116p and other DEAD-box proteins^{10,13}, is formed upon binding ATP and single-stranded RNA (ssRNA) and is thought to represent a ‘post-unwound’ state of the enzyme.

The wide separation of D1 and D2 of Mss116p in the open state suggests that they might function independently to recognize ATP and RNA substrates. To investigate the roles of the individual helicase core domains, we compared the ATP and double-stranded RNA (dsRNA) binding properties of the full core (D1D2) and isolated D1 and D2 of Mss116p. Gel-filtration and ATP–agarose binding assays show that ATP binds to D1 and the full core with similar affinities, but does

not bind appreciably to D2 (Fig. 1b and Supplementary Fig. 1). This result is consistent with previous studies that establish D1 of DEAD-box proteins as a conserved ATP-binding domain^{14,15}. Conversely, fluorescence anisotropy and electrophoretic mobility shift assays (EMSA) show that a 14-base pair (bp) dsRNA binds to D2 and the full core with similar affinities, but does not bind appreciably to D1 (Fig. 1c and Supplementary Fig. 2). The preferential binding of ATP by D1 and dsRNA by D2 with similar affinities to the full core supports the hypothesis that these helicase domains function independently in initial substrate capture in the open state of Mss116p.

We next determined a crystal structure of Mss116p D2 in complex with the same 14-bp dsRNA used in the binding assays (Fig. 2a) at

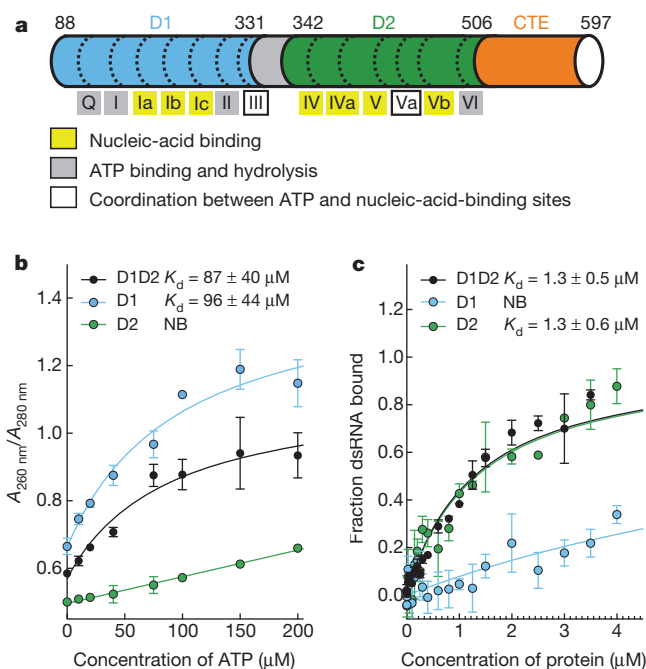


Figure 1 | The distinct substrate-binding characteristics of the helicase core domains of Mss116p. **a**, Schematic of the domain architecture of the helicase core of Mss116p (D1, blue; D2, green; CTE of D2, orange), indicating conserved DEAD-box sequence motifs defined according to ref. 24. Full-length Mss116p contains additional unstructured N-terminal (residues 37–87) and C-terminal (residues 598–664) regions that are not required for helicase activity^{8,9}. **b**, Affinity of ATP for D1, D2, and the full helicase core (D1D2) measured by gel-filtration chromatography under equilibrium conditions. ATP binding was also assessed by an ATP–agarose binding assay (Supplementary Fig. 1). **c**, Affinity of FAM–dsRNA (Fig. 2a) for MBP-tagged D1, D2 and D1D2 determined by fluorescence anisotropy measurements. Similar results for dsRNA binding were obtained by EMSA (Supplementary Fig. 2). Error bars in b and c represent the standard error for at least three independent measurements, and the error in the K_d represents the standard error of the nonlinear regression (see Methods). NB, no significant binding.

¹Institute for Cellular and Molecular Biology, Department of Chemistry and Biochemistry, and Section of Molecular Genetics and Microbiology, School of Biological Sciences, University of Texas at Austin, Austin, Texas 78712, USA. [†]Present address: Rigaku Americas Corporation, 9009 New Trails Drive, The Woodlands, Texas 77381, USA.

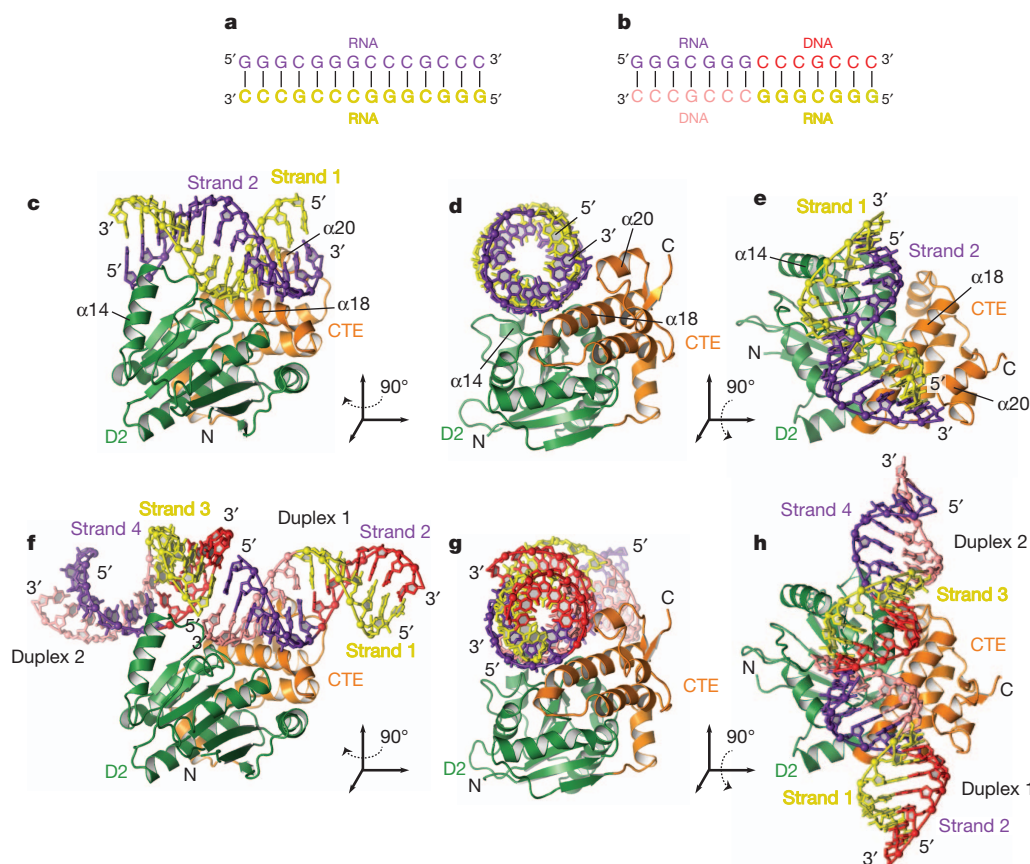


Figure 2 | Crystal structures of Mss116p D2 bound to A-form duplexes.

a, The 14-bp self-complementary GC-rich RNA-duplex substrate. **b**, The 14-bp GC-rich chimaeric RNA-DNA-duplex substrate. **c–e**, Orthogonal views of the D2–dsRNA complex coloured as in **a** and Fig. 1a. Helix $\alpha 14$ of D2, which

contains motif IVa, faces the major groove of the dsRNA, and $\alpha 18$ and $\alpha 20$ of the CTE face the minor groove of the dsRNA. **f–h**, Orthogonal views of the D2–dsRNA–DNA complex, coloured as in **b** and Fig. 1a, in which D2 is bound to two stacked 14-bp chimaeric RNA–DNA duplexes.

3.2 Å resolution, the first structure of a DEAD-box helicase domain bound to a duplex substrate (Fig. 2c–e, Supplementary Fig. 3 and Supplementary Table 1). The crystallographic asymmetric unit contains four very similar complexes with protein molecules bound on either side of a pseudo-continuous RNA duplex (Supplementary Fig. 3a). The structure of a single complex shows that D2 contains a positively charged binding pocket for an RNA duplex of A-form geometry (Fig. 2c–e and Supplementary Fig. 3b). One duplex strand (strand 1) interacts extensively with D2 (Fig. 3a, b and Supplementary Table 2). These interactions include multiple contacts to the phosphate groups of the three centrally bound nucleotide residues (N4–N6) by DEAD-box motifs IV, IVa, V, and a loop containing motif Va. The second strand (strand 2) makes only a few contacts, which include hydrogen bonds between 2'-OH groups and the CTE (Fig. 3a, c and Supplementary Table 2). No protein contacts are observed to the RNA bases of either strand (Fig. 3), consistent with the non-specific RNA binding shown by Mss116p and other DEAD-box proteins¹. Except for the contact by motif Va, the structure of D2 and its contacts to the phosphate backbone of strand 1 are the same as in the closed-state structure of Mss116p, in which D2 additionally interacts with D1 and adenosine nucleotide (Fig. 3a and Supplementary Fig. 4; root mean squared deviation = 0.46 Å)¹⁰. Given the similar binding affinities observed for dsRNA by isolated D2 and the full helicase core (Fig. 1c and Supplementary Fig. 2), we propose that the D2–dsRNA structure provides a model for the initial complex of Mss116p with duplex RNA in the open state of the enzyme and a structural basis for dsRNA recognition by Mss116p.

We also determined a structure of D2 in complex with an equivalent 14-bp chimaeric RNA–DNA duplex (Fig. 2b) at 3.6 Å resolution, in which each molecule of protein interacts with two A-form duplex

substrates (Fig. 2f–h, Supplementary Fig. 5 and Supplementary Table 1). These chimaeric duplexes make almost identical contacts with D2 as dsRNA (Supplementary Fig. 5b). Surprisingly, however, a DNA segment of strand 1 interacts with the DEAD-box motifs in the main 'RNA'-binding tract of D2, while an RNA segment binds to the CTE (Fig. 2f–h and Supplementary Fig. 5). This orientation is probably favoured because D2 interacts primarily with nucleic acid substrate phosphate groups, whereas the non-conserved CTE makes hydrogen-bond contacts with the 2'-OHs of RNA (see earlier). The finding that A-form DNA and RNA interact in a similar manner in the conserved RNA-binding tract of D2 suggests that the substrate specificity of DEAD-box proteins for RNA duplexes is dictated primarily by phosphate backbone geometry. Consistent with this idea, modelling demonstrates that the binding pocket of D2 is not shaped to recognize a B-form DNA duplex, the predominant conformation of dsDNA (Supplementary Fig. 6). Additionally, a genetic assay that is stringently dependent upon Mss116p function indicates that the side chains of conserved residues of D2 that interact with phosphate groups (R415, motif IVa; and T433, motif V) are critical for Mss116p function *in vivo*⁹, whereas the side chains of residues in the CTE that interact with 2'-OH groups of the RNA (S535, R538) can be mutated without detectable loss of function (Supplementary Fig. 7). Nucleic acid recognition based on duplex geometry may explain why DEAD-box proteins can unwind chimaeric RNA–DNA duplexes with as few as two centrally located ribonucleotides³, as structural studies indicate that chimaeric duplexes with only one ribonucleotide can adopt A-form geometry^{16,17}.

Our D2 structures provide insight into several other DEAD-box protein activities. DEAD-box proteins function on a wide variety of RNA substrates^{1,2}. A comparison of the D2–dsRNA and D2–dsRNA–DNA structures, and different complexes within their asymmetric

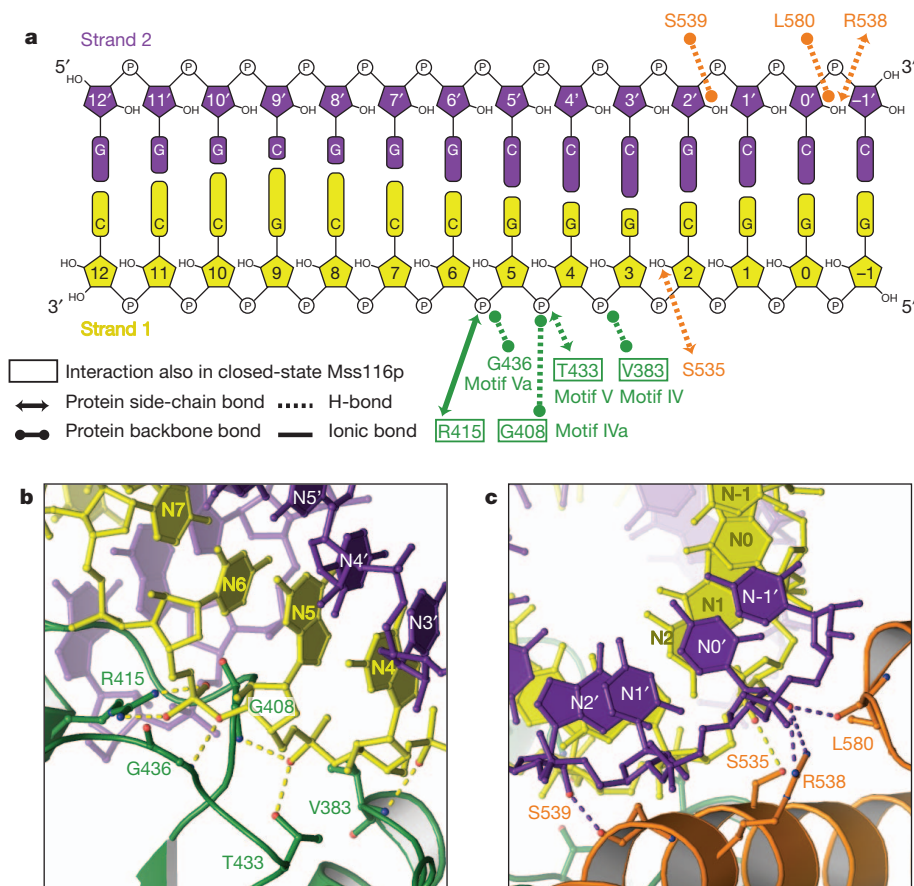


Figure 3 | Interactions between Mss116p D2 and duplex RNA. a, Schematic of RNA–protein interactions observed in the D2–dsRNA structure. The dsRNA interacts with conserved DEAD-box motifs IV–Va of D2 (green) and the CTE of D2 (orange). Boxes indicate that the interaction is maintained in closed-state Mss116p¹⁰. RNA bases are numbered according to their position

units, shows that the binding orientation of the distal regions of duplex substrates can vary, while the contacts between the conserved motifs and centrally bound nucleotide residues are maintained (Supplementary Fig. 8). Although these differences in substrate binding orientation could be influenced by crystal packing, the observed flexibility in nucleic-acid binding away from the basic binding tract of D2 may be advantageous for the loading and unwinding of diverse physiological RNA substrates, such as group I and group II introns, and could contribute towards the general RNA-chaperone activity of Mss116p^{5,6}. The presence of a dsRNA-binding pocket in D2 also raises the possibility that this domain could have a role in strand annealing in the absence of ATP, an activity observed for Mss116p and other DEAD-box proteins¹, by orienting two ssRNAs in a position to pair in the duplex-binding pocket. The additional RNA interactions with the CTE of Mss116p may explain the relatively high strand annealing activity of Mss116p compared to other DEAD-box proteins¹⁸.

Collectively, our results indicate that RNA unwinding by Mss116p begins with the helicase core domains functioning independently to bind ATP and RNA substrates (Fig. 4a). This previously unobserved mechanism for substrate recognition by a helicase is consistent with the wide separation of the two core domains (~ 50 Å for their centres of mass) in the solution structure of the open state¹¹. Subsequent interactions of the exposed regions of these substrates with the remainder of their binding sites in the opposite domain would result in cooperative tight binding coupled to core closure, RNA-strand separation, and formation of the ATPase active site (Fig. 4a). Notably, the loop in D2 that contains part of motif Va and interacts with duplex RNA in the D2-dsRNA structure (see earlier) shifts markedly upon formation

relative to ssRNA in closed-state Mss116p (see Supplementary Fig. 4d). Similar nucleic-acid-protein interactions were observed in the D2-dsRNA-DNA structure (Supplementary Fig. 5). H-bond, hydrogen bond. **b**, Interactions between strand 1 (yellow) of the duplex RNA and D2 (green). **c**, Interactions between duplex RNA and the CTE of D2 (orange).

of the closed state and helps form the ATPase active site (Fig. 4b). This conformational change may be part of a switch that triggers ATP hydrolysis upon core closure. After ATP hydrolysis, dissociation of P_i and ADP cause reopening of the core, release of the bound strand, and regeneration of the enzyme^{8,19}.

Comparison of the D2–dsRNA structure to the previously reported closed-state structure of the helicase core of Mss116p bound to adenosine nucleotide and ssRNA (Fig. 4c) indicates that D2 functions as a stationary platform that positions dsRNA for unwinding by the incursion of D1 (Fig. 4d, e and Supplementary Fig. 9). In the closed state of Mss116p, RNA-binding motifs 1a, 1b, 1c, and the post-II region of D1 are sterically incompatible with dsRNA bound in the duplex-binding site of D2 (Fig. 4d). This suggests that D1 promotes RNA unwinding in two ways (Fig. 4e and Supplementary Fig. 9). First, the conserved post-II region of D1 interrupts the centrally bound base pairs of the RNA duplex to displace strand 2. This displacement is presumably facilitated by the minimal interaction of strand 2 with the protein and could occur actively or during ‘breathing’ of the duplex²⁰. Second, core closure introduces two bends in strand 1, one by interactions with the conserved ‘wedge helix’ in D1 (motif 1c) and the other by interactions with the CTE^{9,10}. DEAD-box proteins that unwind RNA but lack the CTE introduce only the first bend using the same conserved D1 wedge helix¹³. On core closure, the buried solvent accessible surface area of strand 1 increases owing to additional interactions with D1 (1,256 Å² compared to 544 Å² in the D2–dsRNA structure; Supplementary Table 2), contributing to the driving force for RNA unwinding. Further, the two bends induced in strand 1 (Supplementary Fig. 9b) impede its re-annealing to strand 2.

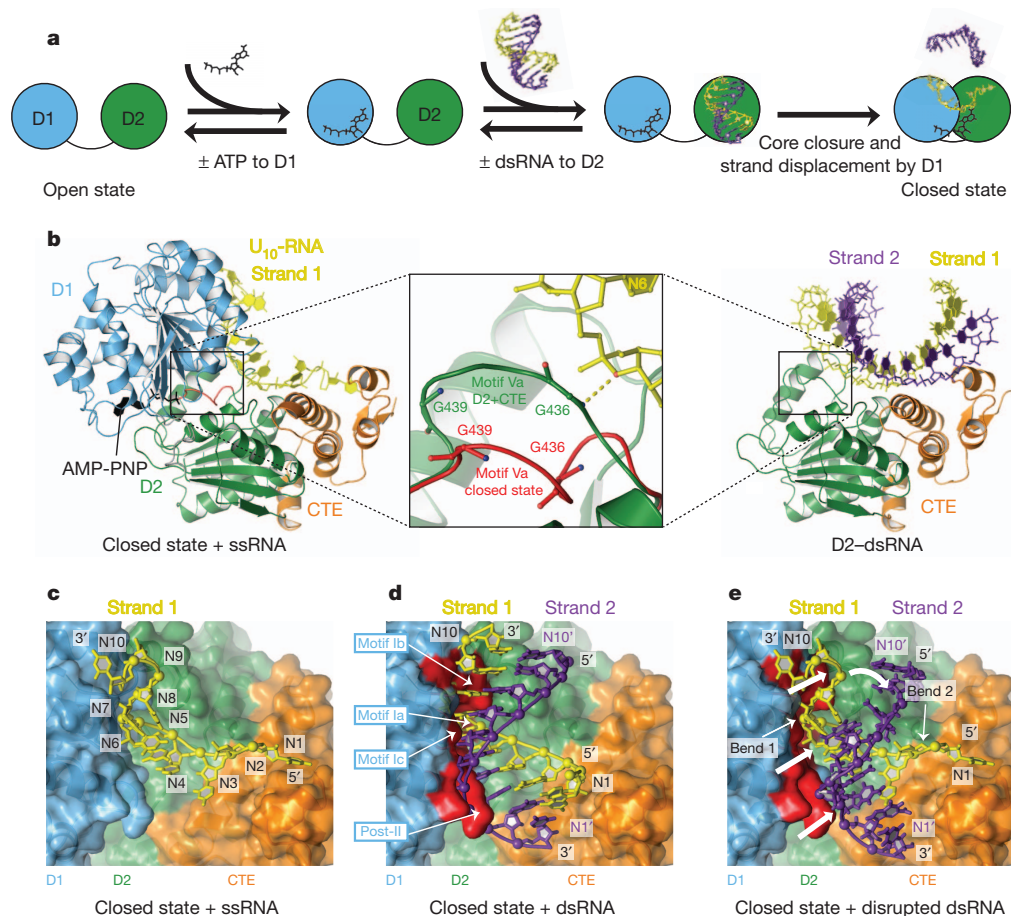


Figure 4 | RNA-duplex binding and unwinding by Mss116p. **a**, Model for the modular roles of the helicase core domains of Mss116p during RNA recognition and unwinding. Although ATP and duplex RNA could bind in either order, the binding of ATP to D1 is shown as a first step because Mss116p-ATP complexes are probably pre-populated at physiological concentrations of ATP (>1 mM)³⁰. **b**, Comparison of the position and interactions of the flexible motif Va loop (residues 435–440) in the D2-dsRNA and closed-state structures of Mss116p. The closed-state helicase core of Mss116p (PDB accession 3I5X)¹⁰ bound to ssRNA (U₁₀-RNA; yellow) and adenosine nucleotide (AMP-PNP; black) is shown with domains coloured as in

Fig. 1a. In the D2-dsRNA structure, the motif Va loop (green) interacts with strand 1 of dsRNA (yellow), whereas in the closed-state structure, the loop (red) shifts to a different position where motif Va helps to form the ATP-binding site. **c**, Surface representation of closed-state Mss116p with N1–N10 of U₁₀-RNA (yellow) indicated. **d**, Surface representation of closed-state Mss116p with dsRNA modelled in the duplex-RNA-binding pocket of D2. Sterically incompatible regions of D1 are highlighted in red. **e**, Change in trajectory of strand 1 and predicted displacement of strand 2 of dsRNA by D1 upon core closure of Mss116p with arrows indicating regions of the substrate that are displaced.

Additional conformational changes that occur upon ATP hydrolysis, dissociation of P_i and/or reopening of the core may also contribute to RNA unwinding.

The model we propose here for RNA-duplex recognition and unwinding by Mss116p explains the previously reported requirement for ATP binding, but not hydrolysis, for RNA unwinding by DEAD-box proteins^{21,22}. This is because substrate binding drives core closure and strand separation (Fig. 4a), whereas ATP hydrolysis to regenerate the enzyme cannot occur until formation of the ATPase active site in the closed-state core. The model also elucidates why RNA unwinding by DEAD-box proteins is non-processive and can initiate directly from a double-stranded region of a substrate^{1,2}, as RNA duplexes are bound directly by D2. Additionally, the differences in RNA-unwinding activity observed for different ATP analogues²² could reflect differences in their binding affinity for the closed state.

Because the RNA-unwinding mechanism described here for Mss116p depends primarily upon conserved DEAD-box protein structures and motifs with only an ancillary role for the CTE, we propose that its major features are used by all DEAD-box proteins. All DEAD-box proteins rely on their helicase core for RNA unwinding and use appended domains for auxiliary functions, such as interactions with partner proteins or to target the helicase core to specific RNA substrates¹. Structural studies of other DEAD-box proteins show that

D1 can by itself bind adenosine nucleotide in a binding pocket formed by the Q-motif, which recognizes the adenine base, motif I (the phosphate-binding or P-loop) and motif II^{14,15}. By contrast, structures of Mss116p and other DEAD-box proteins in the closed state show that D2 interacts minimally with the adenine base^{1,10,13}, in agreement with our observation that D2 does not by itself bind specifically to ATP (Fig. 1b and Supplementary Fig. 1). Likewise, all DEAD-box proteins contain a conserved RNA-binding track in D2 that contains motifs IV, IVa and V¹, and could recognize dsRNA similarly to Mss116p. By contrast, dsRNA is sterically incompatible with the motif Ic wedge helix and the post-II region in the RNA-binding track of D1 of other DEAD-box proteins^{10,13,23}, in agreement with our finding that D1 of Mss116p cannot by itself bind an RNA duplex (Fig. 1c and Supplementary Fig. 2). The post-II region, which displaces strand 2 of a bound duplex in the unwinding mechanism proposed for Mss116p (Fig. 4d, e), is conserved and positioned to have the same role in other DEAD-box proteins^{9,23}, as is motif Va^{1,15,24}, which forms part of loop that contributes to initial binding of the duplex RNA and rearranges to help form the ATPase active site in the closed state (Fig. 4b).

Non-ring-forming helicases with structurally conserved cores of RecA-like domains D1 and D2 are classified into two superfamilies (SFs), SF1 and SF2, with DEAD-box proteins comprising the largest family of SF2 (refs 24, 25). These helicases are thought to have evolved

from a common ancestor²⁶, but have diverged to possess a variety of accessory domains, to have different specificities for RNA or DNA substrates, and to operate by distinct mechanisms^{24,25}. These include processive and non-processive duplex unwinding and translocation without unwinding. The modular substrate-binding functions of D1 and D2, duplex binding by D2, and substrate specificity based on nucleic acid geometry found here for Mss116p may be features that underlie these diverse mechanisms. Interestingly, two recent crystal structures of the closed state of the pathogen recognition receptor RIG-I (also known as DDX58), an SF2 helicase closely related to DEAD-box proteins, show dsRNA bound in the RNA-binding track of D2 of RIG-I in the same orientation as D2 of Mss116p and interacting with conserved SF2 helicase RNA-binding motifs including IV, IVa and V (Supplementary Fig. 10a, c)^{27,28}. However, the orientation of D1 in the RIG-I complex differs from that of Mss116p in the closed state (Supplementary Fig. 10b), enabling RIG-I to bind and translocate on a duplex substrate without RNA unwinding^{27,28}. RIG-I also contains ancillary domains that contribute to dsRNA binding and may influence substrate orientation to favour duplex binding over unwinding^{27–29}. Other SF1 and SF2 helicases could have evolved similarly to promote different closed-state conformations of their helicase cores that give rise to distinct mechanisms of action.

METHODS SUMMARY

The helicase core of Mss116p (D1D2; residues 88–597), D1 (residues 88–330) and D2 (residues 342–597) were expressed and purified as described for full-length Mss116p^{10,11}. Substrate-binding assays were performed in a buffer of 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 10% glycerol, 1 mM dithiothreitol, 5 mM MgCl₂. ATP binding was measured by incubating proteins with ATP at 22 °C followed by gel-filtration chromatography at 4 °C with increasing concentrations of ATP to measure the $A_{260\text{ nm}}/A_{280\text{ nm}}$ of the eluted protein. For ATP-agarose assays, proteins were incubated with ATP-agarose at 4 °C for 24 h, and binding was assessed by SDS-PAGE of the ATP-agarose pellet. Fluorescence anisotropy and EMSA measurements were performed using a 3' fluorescein (FAM)-labelled 14-bp self-complementary dsRNA (5'-GGGCGGGCCCCGCC-FAM-3') annealed by slow cooling after heating to 94 °C for 1 min. Crystals of D2-dsRNA that contain an unlabelled version of the dsRNA substrate used in the binding assays were obtained in 8% tacisamate (pH 6.0; Hampton Research), 20% (w/v) PEG 3350 by hanging drop. Crystals diffracted to 3.2 Å and belong to space group $P2_12_12_1$ with a unit cell of $a = 160.5\text{ Å}$, $b = 88.4\text{ Å}$ and $c = 121.2\text{ Å}$. The final solution has an R_{work} and R_{free} of 22.4% and 26.8%, respectively. Crystals of D2-dsRNA-DNA that contain a 14-bp self-complementary chimaeric substrate (5'-rGrGrGrGrGrGrGdCdCdCdCdCdCdC) were obtained in 6% tacisamate (pH 5.0; Hampton Research), 20% (w/v) PEG 3350 by hanging drop. Crystals diffracted to 3.6 Å and belong to space group $P2_12_12_1$ with a unit cell of $a = 43.7\text{ Å}$, $b = 70.1\text{ Å}$ and $c = 214.9\text{ Å}$. The final solution has an R_{work} and R_{free} of 24.4% and 28.0%, respectively.

Full Methods and any associated references are available in the online version of the paper.

Received 22 May; accepted 6 July 2012.

Published online 2 September 2012.

1. Linder, P. & Jankowsky, E. From unwinding to clamping—the DEAD box RNA helicase family. *Nature Rev. Mol. Cell Biol.* **12**, 505–516 (2011).
2. Jarmoskaite, I. & Russell, R. DEAD-box proteins as RNA helicases and chaperones. *WIREs: RNA* **2**, 135–152 (2011).
3. Yang, Q., Del Campo, M., Lambowitz, A. M. & Jankowsky, E. DEAD-box proteins unwind duplexes by local strand separation. *Mol. Cell* **28**, 253–263 (2007).
4. Pan, C. & Russell, R. Roles of DEAD-box proteins in RNA and RNP folding. *RNA Biol.* **7**, 667–676 (2010).
5. Huang, H. R. *et al.* The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proc. Natl Acad. Sci. USA* **102**, 163–168 (2005).
6. Del Campo, M. *et al.* Unwinding by local strand separation is critical for the function of DEAD-box proteins as RNA chaperones. *J. Mol. Biol.* **389**, 674–693 (2009).
7. Potratz, J. P., Del Campo, M., Wolf, R. Z., Lambowitz, A. M. & Russell, R. ATP-dependent roles of the DEAD-box protein Mss116p in group II intron splicing *in vitro* and *in vivo*. *J. Mol. Biol.* **411**, 661–679 (2011).
8. Cao, W. *et al.* Mechanism of Mss116 ATPase reveals functional diversity of DEAD-box proteins. *J. Mol. Biol.* **409**, 399–414 (2011).

9. Mohr, G. *et al.* High-throughput genetic identification of functionally important regions of the yeast DEAD-box protein Mss116p. *J. Mol. Biol.* **413**, 952–972 (2011).
10. Del Campo, M. & Lambowitz, A. M. Structure of the yeast DEAD-box protein Mss116p reveals two wedges that crimp RNA. *Mol. Cell* **35**, 598–609 (2009).
11. Mallam, A. L. *et al.* Solution structures of DEAD-box RNA chaperones reveal conformational changes and nucleic acid tethering by a basic tail. *Proc. Natl Acad. Sci. USA* **108**, 12254–12259 (2011).
12. Wang, S., Overgaard, M. T., Hu, Y. & McKay, D. B. The *Bacillus subtilis* RNA helicase YxiN is distended in solution. *Biophys. J.* **94**, L01–L03 (2008).
13. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. *Cell* **125**, 287–300 (2006).
14. Rudolph, M. G., Heissmann, R., Wittmann, J. G. & Klostermeier, D. Crystal structure and nucleotide binding of the *Thermus thermophilus* RNA helicase Hera N-terminal domain. *J. Mol. Biol.* **361**, 731–743 (2006).
15. Schütz, P. *et al.* Comparative structural analysis of human DEAD-box RNA helicases. *PLoS ONE* **5**, e12791 (2010).
16. Egli, M., Usman, N. & Rich, A. Conformational influence of the ribose 2'-hydroxyl group: crystal structures of DNA-RNA chimeric duplexes. *Biochemistry* **32**, 3221–3237 (1993).
17. Wahl, M. C. & Sundaralingam, M. B-form to A-form conversion by a 3'-terminal ribose: crystal structure of the chimera d(CCACTAGT)g(G). *Nucleic Acids Res.* **28**, 4356–4363 (2000).
18. Halls, C. *et al.* Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity. *J. Mol. Biol.* **365**, 835–855 (2007).
19. Henn, A. *et al.* Pathway of ATP utilization and duplex rRNA unwinding by the DEAD-box helicase, DbpA. *Proc. Natl Acad. Sci. USA* **107**, 4046–4050 (2010).
20. Rocak, S. & Linder, P. DEAD-box proteins: the driving forces behind RNA metabolism. *Nature Rev. Mol. Cell Biol.* **5**, 232–241 (2004).
21. Chen, Y. *et al.* DEAD-box proteins can completely separate an RNA duplex using a single ATP. *Proc. Natl Acad. Sci. USA* **105**, 20203–20208 (2008).
22. Liu, F., Putnam, A. & Jankowsky, E. ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl Acad. Sci. USA* **105**, 20209–20214 (2008).
23. Andersen, C. B. F. *et al.* Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* **313**, 1968–1972 (2006).
24. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* **20**, 313–324 (2010).
25. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23–50 (2007).
26. Goralenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* **17**, 4713–4730 (1989).
27. Kowalinski, E. *et al.* Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* **147**, 423–435 (2011).
28. Jiang, F. *et al.* Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature* **479**, 423–427 (2011).
29. Luo, D. *et al.* Structural insights into RNA recognition by RIG-I. *Cell* **147**, 409–422 (2011).
30. Ozalp, V. C., Pedersen, T. R., Nielsen, L. J. & Olsen, L. F. Time-resolved measurements of intracellular ATP in the yeast *Saccharomyces cerevisiae* using a new type of nanobiosensor. *J. Biol. Chem.* **285**, 37579–37588 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Monzingo (University of Texas at Austin Macromolecular Crystallography Facility) for help with X-ray diffraction data collection, and R. Russell (University of Texas at Austin) and E. Jankowsky (Case Western Reserve University) for comments on the manuscript. X-ray diffraction data were collected at the Berkeley Center for Structural Biology, which is supported in part by the National Institutes of Health (NIH), National Institute of General Medical Sciences and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract no. DE-AC02-05CH11231. A.L.M. is the recipient of an EMBO long-term fellowship (ALTF 389-2010). This work was supported by NIH grant GM037951.

Author Contributions This study was designed by A.L.M. and A.M.L. A.L.M. cloned the individual domain constructs, purified proteins, crystallized the complexes, collected X-ray crystallographic data, and performed binding assays. A.L.M., M.D.C. and D.J.S. processed and refined the X-ray diffraction data. B.G. performed genetic assays. All authors contributed to analysing the results. A.L.M., M.D.C. and A.M.L. wrote the paper, with contributions from D.J.S. and B.G.

Author Information Coordinates and structure factors were deposited in the Protein Data Bank under accessions 4DB2 (D2-dsRNA) and 4DB4 (D2-dsRNA-DNA). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M.L. (ilambowitz@austin.utexas.edu).

METHODS

Oligonucleotides. The RNA and RNA–DNA oligonucleotides rGrGrCrGrGrGrGrGrCrCrCrGrCrCrC and rGrGrGrGrGrGrGdCdCdCdCdCdCdCdC (Integrated DNA Technologies) were annealed to form 14-bp RNA or chimeric RNA–DNA duplexes by heating 6 mM solutions in 100 mM potassium acetate, 30 mM HEPES (pH 7.5) at 94 °C for 1 min and then slowly cooling to room temperature over 1 h.

Protein expression and purification. pMAL-Mss116p contains the coding sequence for Mss116p (codons 37–664) with an in-frame N-terminal MalE fusion cloned downstream of a *tac* promoter in the expression vector pMAL-c2t (a derivative of pMAL-c2x; New England Biolabs)¹⁸. Mss116p/D1D2 is a derivative of pMAL-Mss116p that expresses the active helicase core of Mss116p (residues 88–597) with deletions of an unstructured N-terminal extension and C-terminal tail¹¹. Expression vectors for the Mss116p constructs D1 (residues 88–330) and D2 (residues 342–597) were created by PCR of pMAL-Mss116p with primers that introduce BamHI and HindIII sites at the 5' and 3' end of the desired gene segment, and then cloning the PCR product between the corresponding sites of pMAL-c2t to link the protein-coding sequence to that of the MalE tag.

Mss116p D1D2, D1 and D2 were expressed as N-terminal MalE fusions in *Escherichia coli* Rosetta 2 (EMD Biosciences), grown in ZYP-5052 auto-inducing medium for 24 h at 22 °C, and purified at 4 °C, as described^{10,11}. Purification steps included: (1) removal of nucleic acids by polyethyleneimine precipitation; (2) isolation of MBP-Mss116p by amylose-affinity chromatography (New England Biolabs); (3) removal of the MBP tag by digestion with tobacco etch virus protease; (4) isolation of Mss116p by heparin-Sepharose chromatography (GE Healthcare); and (5) purification and buffer exchange by gel-filtration chromatography using a Superdex S200 column (GE Healthcare). MBP-tagged proteins were purified in an identical manner but without step (3). Proteins were concentrated for crystallization by using a 10-kDa MWCO concentrator (Millipore), and protein concentrations were determined by Bradford Assay (Bio-Rad). Crystallization and storage buffers were 20 mM Tris-HCl (pH 7.5), 200 mM KCl, 10% glycerol, 1 mM dithiothreitol. **Crystallization.** For the D2–dsRNA complex, the protein (~500 µM) was incubated with dsRNA (650 µM duplex) and MgCl₂ (2 mM) for 30 min on the desktop. Hanging drops were assembled using 1 µl of complex and 1 µl of a well solution of 8% tacsimate (pH 6.0; Hampton Research), 20% PEG 3350. Drops were stored at 22 °C and plate-like crystals appeared within 1 week. Crystals were stabilized in a cryoprotectant containing the crystallization solution plus 20% glycerol before flash cooling in liquid N₂. Crystals of D2–dsRNA–DNA were obtained similarly with hanging drops assembled from 1 µl of dsRNA–DNA complex and 1 µl of a well solution of 6% tacsimate (pH 5.0; Hampton Research), 20% PEG 3350 and were cryoprotected as above.

Structure determination. X-ray diffraction data were collected either on our in-house system (Rigaku MicroMax-007 HF generator with VariMax HF optics and an R-Axis IV++ imaging plate detector; wavelength 1.54178 Å) or at the Advanced Light Source (ALS), Lawrence Berkeley National Laboratory (mail-in service on beamlines 5.0.2, or 5.0.3; wavelength 1.00003 Å). Details of data collection and refinement are in Supplementary Table 1. Diffraction intensities were indexed, integrated, and scaled with HKL-2000³¹ or autoPROC³². For diffraction data processed with HKL-2000, additional statistics were calculated from unmerged data using d*TREK³³. Initial space groups were determined by using Pointless³⁴ and confirmed by decreases in both R_{work} and R_{free} after refinement of molecular replacement solutions. Molecular replacement was performed with Phaser³⁵, using the previously determined structure of Mss116p D2 in the closed state (PDB accession 3I5X) as a search model. Composite omit maps were calculated to determine that there was no model bias³⁶. Structures were completed with cycles of manual model building in Coot³⁷ and refinement in Phenix³⁶. Validation of protein and nucleic acid models and their contacts was done by using MolProbity³⁸ and indicated that at least 96% of residues are located in the most favourable region of the Ramachandran plot. Structural figures were prepared by using the PyMOL Molecular Graphics System, v. 1.4. SASA calculations and interface analyses were performed using PDBePISA³⁹.

ATP-binding assays. Equilibrium binding of ATP to D1, D2 and D1D2 was measured by gel filtration and ATP-agarose binding assays. Because the helicase core of Mss116p does not contain tryptophan residues, its calculated extinction coefficient is small ($\epsilon_{280} = 18,255 \text{ M}^{-1} \text{ cm}^{-1}$; ExPASy Proteomics Server ProtParam tool⁴⁰). The binding of ATP ($\epsilon_{260} \sim 15,400 \text{ M}^{-1} \text{ cm}^{-1}$) therefore gives rise to a large change in $A_{260 \text{ nm}}$ compared to $A_{280 \text{ nm}}$. Protein samples (10 µM) were incubated at 22 °C for 30 min in increasing concentrations of ATP-Mg²⁺ (0–200 µM) and loaded onto a HiTrap desalting column (GE Healthcare) pre-equilibrated in a buffer containing the same amount of ATP-Mg²⁺ and

20 mM Tris-HCl (pH 7.5), 100 mM KCl, 10% glycerol, 1 mM dithiothreitol, 5 mM MgCl₂. The absorbance of the eluted protein above the background signal of the buffer was measured at 260 and 280 nm, and the change in $A_{260 \text{ nm}}/A_{280 \text{ nm}}$ signal at increasing concentrations of ATP was fit to a simple one-site ligand-binding model to calculate a K_d for the protein–ATP complex.

ATP-agarose binding assays were performed in 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 10% glycerol, 1 mM dithiothreitol, 5 mM MgCl₂ by incubating AP-ATP-agarose (150 µl; Jena Bioscience) with ~50 µg of protein overnight at 4 °C with agitation. The beads were pelleted by centrifugation (1,000g, 1 min), the supernatant was removed, and the pellet was resuspended in binding buffer (150 µl). The load, supernatant, and ATP-agarose pellet were analysed by SDS-PAGE, and the gels were stained with SYPRO Ruby protein gel stain (Invitrogen) to detect and quantify bound and unbound Mss116p.

RNA-binding assays. Equilibrium binding of dsRNA to D1, D2 and D1D2 was measured by fluorescence anisotropy and EMSA assays. Fluorescence anisotropy measurements were performed using MBP-tagged proteins to increase the change in anisotropy signal upon binding. A 3' FAM-labelled 14-bp RNA duplex (10 nM; IDT; Fig. 2a) was incubated with increasing concentrations of protein (0–4 µM) at 22 °C for at least 1 h in a reaction medium containing 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 10% glycerol, 1 mM dithiothreitol, 5 mM MgCl₂ and 0.1 mg ml⁻¹ of bovine serum albumin to stabilize the protein at low concentrations. The observed fluorescence anisotropy, r_{obs} of FAM-dsRNA at increasing concentrations of protein was measured by using an EnVision Microplate Reader (Perkin Elmer) and was fit to the equation:

$$r_{\text{obs}} = r_{\text{fb}} + r_{\text{f}}(1 - f_{\text{b}})$$

where f_{b} is the fraction of protein-bound dsRNA and r_{f} and r_{b} are the anisotropy of the free and protein-bound dsRNA, respectively⁴¹. f_{b} was defined as $f_{\text{b}} = [RP]/[P]$, where $[RP]$ and $[P]$ are the concentration of protein–RNA complex and the total protein concentration, respectively. The dissociation constant, K_d , is defined by the quadratic equation⁴¹:

$$[RP] = \frac{(K_d + [P] + [R_t]) - \sqrt{(K_d + [P] + [R_t])^2 - 4[P][R_t]}}{2}$$

where $[R_t]$ is the total concentration of dsRNA.

EMSA measurements were performed on both MBP-tagged proteins (to increase protein solubility under the experimental conditions) and untagged proteins. A 3' FAM-labelled 14-bp RNA duplex (100 nM; IDT; Fig. 2a) was incubated with increasing concentrations of protein (0–3 µM) at 22 °C as described above. Samples were then analysed in a non-denaturing 6% polyacrylamide gel run at 4 °C for 60 min, and the fluorescence signal of the bound duplex substrate was quantified on a Typhoon imager (GE Healthcare). Control gels verified that the fluorescence signal of the unbound substrate increased linearly with concentration (Supplementary Fig. 2e).

Genetic assays. Yeast genetic selections of functional Mss116p variants from centromere-containing (CEN) plasmid libraries in which specified codons were randomized and glycerol growth tests of individual variants were performed as described⁹.

- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D* **67**, 293–302 (2011).
- Pflugrath, J. W. The finer things in X-ray diffraction data collection. *Acta Crystallogr. D* **55**, 1718–1725 (1999).
- Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
- Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
- Krisinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
- Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**, 531–552 (1999).
- Tang, G. Q., Bandwar, R. P. & Patel, S. S. Extended upstream A-T sequence increases T7 promoter strength. *J. Biol. Chem.* **280**, 40707–40713 (2005).

Alternating-access mechanism in conformationally asymmetric trimers of the betaine transporter BetP

Camilo Perez^{1*}, Caroline Koshy^{1*}, Özkan Yildiz¹ & Christine Ziegler^{1,2}

Betaine and Na⁺ symport has been extensively studied in the osmotically regulated transporter BetP from *Corynebacterium glutamicum*, a member of the betaine/choline/carnitine transporter family, which shares the conserved LeuT-like fold of two inverted structural repeats¹. BetP adjusts its transport activity by sensing the cytoplasmic K⁺ concentration as a measure for hyperosmotic stress via the osmosensing carboxy-terminal domain^{2,3}. BetP needs to be in a trimeric state for communication between individual protomers through several intratrimeric interaction sites⁴. Recently, crystal structures of inward-facing BetP trimers have contributed to our understanding of activity regulation on a molecular level^{5,6}. Here we report new crystal structures, which reveal two conformationally asymmetric BetP trimers⁷, capturing among them three distinct transport states. We observe a total of four new conformations at once: an outward-open apo and an outward-occluded apo state, and two closed transition states—one in complex with betaine and one substrate-free. On the basis of these new structures, we identified local and global conformational changes in BetP that underlie the molecular transport mechanism, which partially resemble structural changes observed in other sodium-coupled LeuT-like fold transporters, but show differences we attribute to the osmolytic nature of betaine, the exclusive substrate specificity and the regulatory properties of BetP.

Crystals of a surface-engineered, fully functional BetP mutant (BetP(ΔN29/E44E45E46/AAA))⁵ were grown in the presence of betaine, diffracting to a resolution of 3.1 Å. Exchange of Gly 153 against aspartate in the unfolded stretch of transmembrane helix 1 (TMH)1' in the same mutant (Supplementary Fig. 1) resulted in a sixfold increased affinity for sodium and additional specificity for choline⁶ (Supplementary Fig. 2). Note that the numbering of the BetP TMHs was adapted to the LeuT numbering for better comparison⁸. Therefore, TMH1'–TMH10' correspond to TMH3–TMH12, whereas TMH1 and TMH2 are now designated as TMH(–2) and TMH(–1), respectively. BetP(G153D) was co-crystallized with choline and crystals diffracted to a resolution of 3.25 Å. The structures revealed BetP trimers that lack a non-crystallographic three-fold symmetry (Fig. 1, Supplementary Fig. 3a and Supplementary Table 1). That is, each protomer within one trimer adopts a different conformation of the alternating-access cycle (Supplementary Fig. 3b): a substrate-free outward-occluded (C_{eooc}), a substrate-free outward-open (C_e), a closed substrate-free (C_c), a closed substrate-bound (C_{cS}), an inward-open (C_{iS}) betaine-bound and an inward-open choline-bound (Fig. 1). The closed states reported here for the first time—to our knowledge—for a LeuT-like fold transporter are an intermediate between outward- and inward-facing states (Supplementary Fig. 3b). C_{cS} is characterized by a central binding site (S1 site) that is closed by nearly 14 Å of protein bulk from either side of the membrane. The trimethylammonium group of betaine in the S1 site forms cation–π interactions with Trp 373, Trp 374 and Trp 377, which are arranged in a tryptophan prism (Fig. 1, inset). These residues are

located entirely in TMH6' and comprise the signature motif of the betaine/choline/carnitine transporter (BCCT) family¹. We also identified a positive peak in the $F_o - F_c$ difference electron density map in the C_{cS} state localized at the structurally conserved Na2 site (Khafizov *et al.*, manuscript in preparation). Sodium is coordinated in the Na2 site by carbonyls from the backbone of residues Ala 147 and Met 150 from TMH1' and Phe 464 as well as by the hydroxyl groups of residues Thr 467 and Ser 468 in TMH8' (Fig. 1, inset). The betaine location observed in the S1 site is different from that in C_{iS} (Fig. 1, inset) or the published inward-occluded state (C_{ioocS})⁵ (Supplementary Fig. 4). In the C_{ioocS} state, Trp 374 in TMH6' and Trp 194 and Tyr 197 in TMH2'

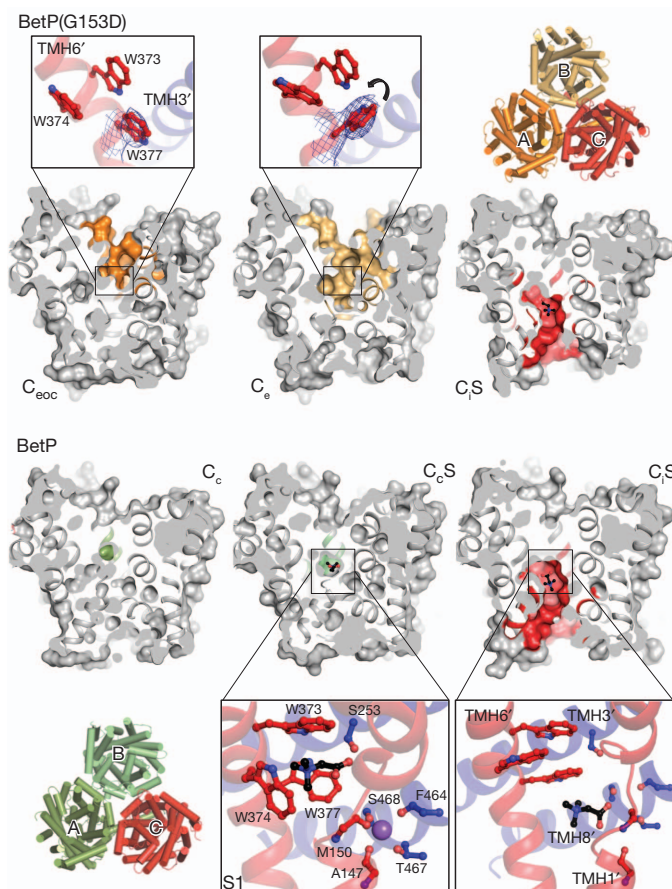


Figure 1 | Conformational states and substrate-binding sites observed in BetP and BetP(G153D) asymmetric trimers. BetP(G153D) (PDB accession 4DOJ): chain A, C_{eooc}; chain B, C_e; chain C, C_{cS}. BetP (PDB accession 4AIN): chain A, C_e; chain B, C_{cS}; chain C, C_{iS}. Betaine (4AIN) and choline (4DOJ) are shown in blue, red and black; sodium is shown in purple. S1 is the central betaine-binding site. The $2F_o - F_c$ map of Trp 377 in the C_e and C_{eooc} states is presented at the 1.4σ level.

¹Max-Planck Institute of Biophysics, Department of Structural Biology, D-60438 Frankfurt am Main, Germany. ²Institute of Biophysics and Biophysical Chemistry, University of Regensburg, D-93053 Regensburg, Germany.

*These authors contributed equally to this work.

coordinate betaine⁵, whereas in C_iS , betaine has shifted towards the cytoplasm so that it solely interacts with Trp 377 in TMH6' and backbone carbonyls from TMH1'.

Rotamer states of Trp 374 and Trp 377 change during state transitions, whereas Trp 373 remains in the same plane although with different inclinations of the indole group. Both Trp 374 and Trp 377 rotate by nearly 90°. It is important to note that the binding site in the C_iS and $C_{ioc}S$ states does not form optimal interactions with the substrate; that is, betaine 'fits' perfectly in the S1 site only in the transient C_cS state. Consequently, the binding energy is exploited fully only in that transition state. Trp 377 has a dominant role and any exchange against another residue renders BetP inactive⁵, whereas substitution of Trp 373 or Trp 374 decreases affinity (Supplementary Fig. 5a, b). Side-chain rotations were also observed for several aromatic residues in TMH1' and TMH6' (Phe 156, Phe 369, Phe 380 and Phe 384), which are highly conserved in the BCCT family (Supplementary Fig. 5c). Alanine substitution of any periplasmic occluding residue (Supplementary Fig. 5a, b and Supplementary Table 2) decreases affinity for betaine, whereas mutation of aromatic residues in the cytoplasmic pathway (Supplementary Fig. 5d, e and Supplementary Table 2) did not have a major effect on affinity, suggesting that only the periplasmic aromatic residues directly or indirectly contribute to betaine recruitment. Consequently, betaine is preferably transported from outside to inside. The transition from inward to outward also occurs via a closed, albeit substrate-free C_c state. The closed states C_c and C_cS are very similar to one another (root mean squared deviation 1.0 Å) and the tryptophan prism remains nearly unchanged (Supplementary Fig. 6). As this conformational change occurs in the absence of substrate, a relatively flat free energy landscape might be required to allow BetP to cycle back to an outward-facing state, which is also presented here for the first time for a BCCT. Both outward-facing states are occluded from the cytoplasmic side by ~16 Å of protein bulk (Fig. 1). In the C_e state, a periplasmic funnel extends to the Na2 site (Supplementary Fig. 7a). In the C_{eoc} state, the periplasmic funnel is less deep and the Na2 site is not

accessible, mainly due to the rotation of Trp 377 by nearly 90° (Fig. 1, inset, and Supplementary Fig. 7a). The aromatic side chains described above form the floor in both outward-facing conformations, akin to the outward-open state of the arginine/agmatine antiporter AdiC⁸. C_{eoc} and C_e states show a very similar main-chain conformation, although a 2.0 Å displacement of the periplasmic half of TMH10' facilitates opening of the external cavity (Supplementary Fig. 7b). The transition from the occluded to the open state might occur in BetP in the absence of substrate by stochastic thermodynamic fluctuations, after which sodium binding might trigger the full opening of the transporter to the periplasmic side to allow subsequent binding of betaine.

On the basis of the three major conformations—outward-facing, closed and inward-facing—the alternating-access mechanism in BetP can be described as a hybrid of rigid body movements (Fig. 2) and individual flexing of symmetry-related helices (Figs 2 and 3). The rigid body movement involves the scaffold motif (TMH3' and TMH4' in repeat 1 and TMH8' and TMH9' in repeat 2) tilting about 13° away from an axis running through the centre of BetP normal to the plane of the membrane, relative to the 4-TMH bundle. Similar rigid body rotation has been observed for LeuT⁹ and Mhp1¹⁰, for which the scaffolds tilt by ~10° and ~18°, respectively. However, compared to the tilting of the bundle domain observed for Mhp1¹⁰ and LeuT⁹ this movement is less pronounced in BetP (Supplementary Fig. 8). Recently, we reported two regulatory ionic interaction networks in BetP that are involved in a crosstalk between individual protomers within the trimer¹¹. These networks link bundle helices together, restricting their flexibility, a unique feature only observed for the osmoregulated BetP. We suggest that the restricted opening of BetP is related to its exclusive specificity for betaine, whereas other betaine BCCTs of the OpuD type show low affinity for other osmolytes, for example, proline¹². Betaine is a rather small molecule that acts as an osmolyte and protein stabilizer, promoting folding by forcing proteins to decrease its exposure to solvent by markedly decreasing the water activity¹³. Betaine is excluded from the first hydration shell of the protein backbone; however, in

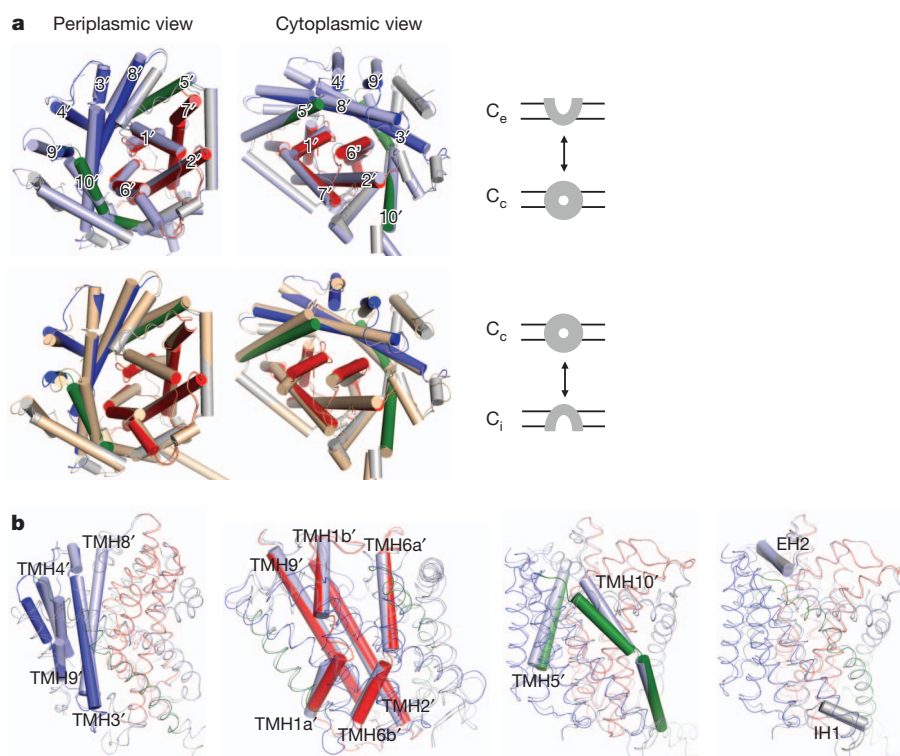


Figure 2 | Conformational changes during the alternating-access cycle.

a, Periplasmic and cytoplasmic views of conformational changes during the transition from the C_e (bundle domain in red (TMHs 1', 2', 6' and 7'), hash motif in blue (TMHs 3', 4', 8' and 9') and thin gates in green (TMHs 5' and

10'); see Supplementary Fig. 1) to the C_c state (light blue), and from the C_c to the C_i state (light yellow). **b**, Superpositions of the C_e (coloured as in Supplementary Fig. 1) and C_i state (light blue) showing in cylinders the scaffold domain, the bundle domain, the 'thin' gates and the extra helical segments, respectively.

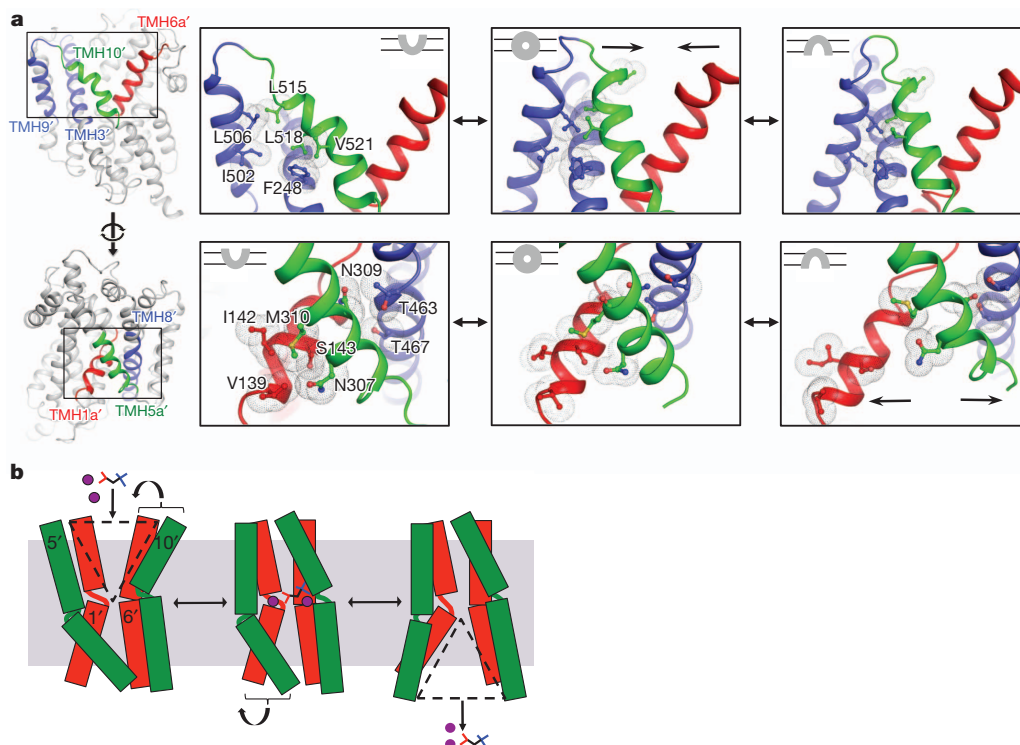


Figure 3 | Opening and closing of the periplasmic and cytoplasmic gates in BetP. **a**, Sequence of opening and closing of gating main chains on both sides of the membrane. From left to right, C_o , C_c and C_i states. **b**, Model of alternating-access 'gating' mechanism in BetP. From left to right, schematic drawings of the C_o , C_c and C_i states. Betaine is shown in black, blue and red, sodium ions in

purple. Arrows indicate the closing of the periplasmic gates TMH6a' and TMH10' after substrate and ions bind to the C_c state, and the opening of the cytoplasmic gates TMH1a' and TMH5' to allow the release of substrate and ions from the C_c state. TMHs are coloured as in Fig. 2 and Supplementary Fig. 1.

transporters interactions with substrates in the micromolar affinity range have to take place¹. From this point of view a fully hydrated substrate pathway owing to a wide opening might be disadvantageous for the stability of BetP and efficient betaine import.

To open the pathways adequately, BetP additionally requires a gating mechanism, observed by the movement of individual helical parts that are not restricted by the regulatory networks (Figs 2 and 3). Similar to what was described recently for LeuT⁹, TMH1a' is displaced considerably to open the cytoplasmic gate. Independently, the symmetry-related TMH6a' contributes to open the periplasmic gate, although with a less pronounced movement. During transition from the outward to inward state, TMH1a' is displaced by ~ 5 Å and tilted by 18° . In LeuT TMH1a is displaced by ~ 12 Å and tilted by $\sim 45^\circ$, which makes this movement the most notable conformational change during the inward-to-outward cycle⁹, positioning TMH1a into the hydrophobic bilayer. The movement in TMH1a' is much smaller in BetP and not even existent in Mhp1. Although flexibility of TMH1a was suggested by single-molecule fluorescence resonance energy transfer (smFRET) studies¹⁴, it is unclear whether the marked conformational change of TMH1a in LeuT is a consequence of the mutations introduced in the crystallization variant causing the weakening of the Na2 binding site and the perturbation of the intracellular gate, and to what extent this movement would take place in the native membrane environment. In the same context, we cannot entirely rule out that in a more native environment BetP might adopt a conformation which exhibits more pronounced flexing of TMH1a'.

The discrepancy between the hinge-like bending movements of TMH1 and TMH6 in LeuT and BetP might also originate from the fact that in LeuT both helices possess a very flexible stretch, whereas in BetP this is true only for TMH1'. The midsection of TMH6', being composed largely of hydrophobic residues (Ser-Pro-Phe-Val), is more rigid, as reflected by the smaller bending angle. On the other hand, hinge-like bending motions of TMH1' were not observed for Mhp1¹⁰

(Supplementary Fig. 8), as the corresponding region (Ala-Ile-Gln-Val-Ala) is quite hydrophobic and presumably rather rigid.

Additional gating-like movements in BetP are observed for TMH5' and TMH10' (Figs 2 and 3), which are assigned as thin gates in the alternating-access mechanism of Mhp1¹⁰. In BetP, TMH5' constrains the cytoplasmic pathway in the outward-facing open state. During substrate translocation it moves out of the central pathway, undergoing a 9° tilt. The structurally equivalent extracellular part of TMH10' constrains the periplasmic pathway in the inward-open state, which is again released when BetP returns to the outward-open state by a movement of 5.7 Å and a massive tilt of 41° relative to the membrane normal. In contrast to the conformational change of TMH5', TMH10' is able to move further away, as it is stabilized by the presence of TMH(-1) and TMH(-2). Interestingly, the presence of a closed transition state implies that these periplasmic and cytoplasmic gates undergo uncoupled hinge movements in BetP.

A new property revealed is the spring-like movement of the unfolded region of TMH1'. This stretch shows remarkable plasticity in BetP, which is related to the transient coordination of sodium ions during the alternating-access cycle (Fig. 4). In fact, the unfolded stretch in TMH1' provides backbone carbonyls for the coordination of sodium in the Na2 site (Fig. 4). The plasticity is therefore affected by binding and dissipation of the positive charge of the coupling sodium in the Na2 site, inducing considerable backbone dihedral changes, as proposed for LeuT^{9,15}, ApcT¹⁵ and vSGLT¹⁶ transporters. In general, partly unfolded transmembrane helices provide the structural flexibility to couple electrochemically favoured binding and release of ions to alternating opening and closing of 'gates'¹⁷. Indeed, all the structural elements involved in the transformation of BetP from the outward- to inward-facing state are either directly involved in the formation of the sodium-binding sites, such as in the case of the bundle helices TMH1' and TMH6' and the hash-motif helices TMH3' and TMH8' (Fig. 4), or indirectly involved such as the thin-gate helices TMH5' and TMH10'

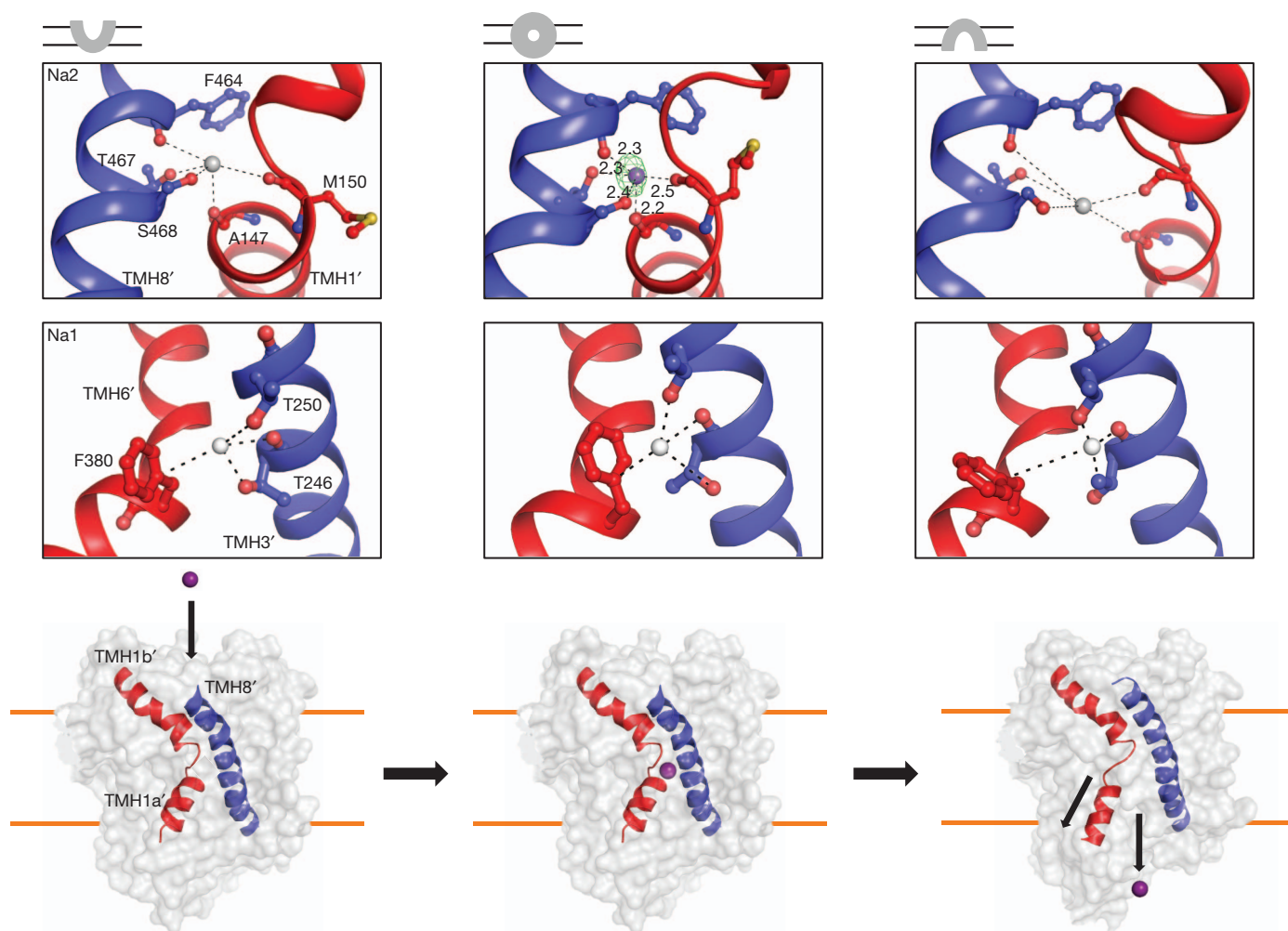


Figure 4 | Conformational changes of the sodium-binding sites.

a, Comparison of the Na2 and Na1 sodium-binding sites in the C_e , C_o and C_i states of BetP. Sodium ions, in white, are presented to schematize the location of the Na2 and Na1 sites. $F_o - F_c$ map, in green, is presented at the 3.0σ level for

the Na2 site located in PDB accession 4AIN. A schematic of sodium-induced conformational change of TMH1a' is also shown. Vectorial movement of sodium through the Na2 site induces a spring-like movement of the cytoplasmic part of TMH1', which renders the cytoplasmic pathway accessible.

(Fig. 3). This correlates well with the absolute requirement of the presence of sodium to allow betaine symport.

In the outward-open and closed states, residues forming the Na2 site provide reasonable coordination for sodium. By contrast, in the inward-open state the coordinating residues are too distant to coordinate a sodium ion, which is a consequence of the main-chain conformational changes of TMH1a' (Fig. 4). We suggest that the Na2 site is perturbed in the inward-open state, which is in agreement with reports for LeuT⁹ and by molecular dynamics (MD) simulations on several LeuT-like fold transporters^{15,16}. Recently, the position of the Na1 site in BetP was identified (Khafizov *et al.*, manuscript in preparation), formed by side chains of residues Phe 380 in TMH6', Thr 246 and Thr 250 in TMH3'. A proper coordination sphere of sodium in the Na1 site is only established in the closed state; in the outward- and inward-facing states residues in TMH3' and TMH6' are too distant to coordinate sodium ions (Fig. 4).

Both ions bind at the interface between the scaffold and bundle helices (Supplementary Fig. 9), which move relative to one another during the alternating-access cycle. The cytoplasmic and periplasmic gates show a symmetry-related mechanism related to the binding and release of sodium ions at the two binding sites. In BetP, sodium binding to Na2 relates to 'opening-closing' of the cytoplasmic gates, while binding to Na1 relates to 'opening-closing' of the periplasmic gates (Figs 3 and 4). In the outward-facing conformation, TMH5', which is

the cytoplasmic 'thin' gate, interacts closely with TMH1a' via a dense network of hydrophilic and hydrophobic interactions. During transition to the inward-open state, release of sodium from the Na2 site triggers the movement of TMH1a', which detaches from TMH5', leading to a full opening of the cytoplasmic pathway. Also, in the outward-facing conformation TMH3' and TMH6a', which coordinate the sodium ion at Na1, are far away from one another, not allowing proper coordination of sodium. In this conformation TMH10', which is the periplasmic 'thin' gate, adopts a conformation that renders the periplasmic pathway accessible, being pulled by interactions with TMH9'. We suggest that binding of sodium to the Na1 site brings together TMH3', TMH6' and TMH9', initiating a cascade of structural rearrangements that result in closure of the periplasmic pathway.

Our data indicate that differences such as the location of sodium-binding sites, the unique coordination of the substrate by the Na1 sodium ion in LeuT¹⁸, or the absence of a Na1 sodium ion as in the case of Mhp1 are responsible for the relative contributions of rigid body rotation versus gating hinge movements observed in each transporter. Consequently, the conservation of the Na2 site is the unifying element in the LeuT-like fold transport mechanism. The alternating-access mechanisms of LeuT-fold transporters share common mechanistic principles; however, for each individual transporter the primary structure and specific coordination of substrate and co-substrates will dictate to what extent 'rocking bundle' and 'gating' can occur.

METHODS SUMMARY

Escherichia coli DH5 α was used for the heterologous expression of *strep-betP*. Membranes were solubilized using β -dodecyl-maltoside (DDM), and BetP was purified by affinity chromatography via Strep-Tactin macroprep and size-exclusion chromatography^{5,19}. Uptake of labelled betaine was measured in *E. coli* MKH13 cells, started by adding 5–250 μ M of [¹⁴C] betaine upon an osmotic shock adjusted with KCl²⁰. Binding assays were performed in proteoliposomes using tryptophan fluorescence emission between 315 and 370 nm with an excitation wavelength set to 295 nm²¹. Data from BetP(Δ N29/E44E45E46/AAA/G153D) crystals obtained in the presence of 1 mM choline were collected to 3.2 Å at beamline id29 at the European Synchrotron Radiation Facility (ESRF). The structure was determined by molecular replacement using BetP (PDB accession 3P03) as the search model (Supplementary Table 1). BetP(Δ N29/E44E45E46/AAA) was crystallized in the presence of 5 mM betaine and a data set to 3.1 Å was collected at the PXII beamline of the Swiss Light Source (SLS). The structure was determined by molecular replacement using BetP (PDB accession 4AMR) as the search model (Supplementary Table 1). For more details, see Supplementary Methods.

Received 18 February; accepted 6 July 2012.

Published online 2 September 2012.

1. Ziegler, C., Bremer, E. & Kramer, R. The BCCT family of carriers: from physiology to crystal structure. *Mol. Microbiol.* **78**, 13–34 (2010).
2. Schiller, D., Rubenhagen, R., Kramer, R. & Mörbach, S. The C-terminal domain of the betaine carrier BetP of *Corynebacterium glutamicum* is directly involved in sensing K⁺ as an osmotic stimulus. *Biochemistry* **43**, 5583–5591 (2004).
3. Rubenhagen, R., Mörbach, S. & Kramer, R. The osmoreactive betaine carrier BetP from *Corynebacterium glutamicum* is a sensor for cytoplasmic K⁺. *EMBO J.* **20**, 5412–5420 (2001).
4. Perez, C., Khafizov, K., Forrest, L. R., Kramer, R. & Ziegler, C. The role of trimerization in the osmoregulated betaine transporter BetP. *EMBO Rep.* **12**, 804–810 (2011).
5. Ressel, S., Terwisscha van Scheltinga, A. C., Vorrhein, C., Ott, V. & Ziegler, C. Molecular basis of transport and regulation in the Na⁺/betaine symporter BetP. *Nature* **458**, 47–52 (2009).
6. Perez, C. *et al.* Substrate specificity and ion coupling in the Na⁺/betaine symporter BetP. *EMBO J.* **30**, 1221–1229 (2011).
7. Tsai, C. J. *et al.* Structural asymmetry in a trimeric Na⁺/betaine symporter, BetP, from *Corynebacterium glutamicum*. *J. Mol. Biol.* **407**, 368–381 (2011).
8. Fang, Y. *et al.* Structure of a prokaryotic virtual proton pump at 3.2 Å resolution. *Nature* **460**, 1040–1043 (2009).
9. Krishnamurthy, H. & Gouaux, E. X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature* **481**, 469–474 (2012).
10. Shimamura, T. *et al.* Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science* **328**, 470–473 (2010).

11. Gärtner, R. M., Perez, C., Koshy, C. & Ziegler, C. Role of bundle helices in a regulatory crosstalk in the trimeric betaine transporter BetP. *J. Mol. Biol.* **414**, 327–336 (2011).
12. Wetzel, K. J., Bjorge, D. & Schwan, W. R. Mutational and transcriptional analyses of the *Staphylococcus aureus* low-affinity proline transporter OpuD during *in vitro* growth and infection of murine tissues. *FEMS Immunol. Med. Microbiol.* **61**, 346–355 (2011).
13. Bourot, S. *et al.* Glycine betaine-assisted protein folding in a lysA mutant of *Escherichia coli*. *J. Biol. Chem.* **275**, 1050–1056 (2000).
14. Zhao, Y. *et al.* Single-molecule dynamics of gating in a neurotransmitter transporter homologue. *Nature* **465**, 188–193 (2010).
15. Shi, L. & Weinstein, H. Conformational rearrangements to the intracellular open states of the LeuT and ApcT transporters are modulated by common mechanisms. *Biophys. J.* **99**, L103–L105 (2010).
16. Watanabe, A. *et al.* The mechanism of sodium and substrate release from the binding pocket of vSGLT. *Nature* **468**, 988–991 (2010).
17. Hunte, C. *et al.* Structure of a Na⁺/H⁺ antiporter and insights into mechanism of action and regulation by pH. *Nature* **435**, 1197–1202 (2005).
18. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na⁺/Cl[−]-dependent neurotransmitter transporters. *Nature* **437**, 215–223 (2005).
19. Ziegler, C. *et al.* Projection structure and oligomeric state of the osmoregulated sodium/glycine betaine symporter BetP of *Corynebacterium glutamicum*. *J. Mol. Biol.* **337**, 1137–1147 (2004).
20. Ott, V., Koch, J., Spate, K., Mörbach, S. & Kramer, R. Regulatory properties and interaction of the C- and N-terminal domains of BetP, an osmoregulated betaine transporter from *Corynebacterium glutamicum*. *Biochemistry* **47**, 12208–12218 (2008).
21. Ge, L., Perez, C., Waclawska, I., Ziegler, C. & Müller, D. J. Locating an extracellular K⁺-dependent interaction site that modulates betaine-binding of the Na⁺-coupled betaine symporter BetP. *Proc. Natl Acad. Sci. USA* **108**, E890–E898 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Forrest and R. Krämer for careful reading of the manuscript and suggestions and discussions. This work was supported by the International Max-Planck Research School (C.K. and C.P.), and by the DFG (German Research Foundation), Collaborative Research Center 807 “Transport and Communication across Biological Membranes” (C.Z.).

Author Contributions C.P. performed mutations, activity and binding measurements in cells and proteoliposomes, crystallization, collection and processing of data for PDB accession 4DOJ; C.K. performed crystallization for PDB accession 4AIN; C.K. and Ö.Y. performed collection and processing of data for PDB accession 4AIN; C.P., C.K. and C.Z. analysed the data; C.Z. directed the research; and C.P. and C.Z. wrote the manuscript.

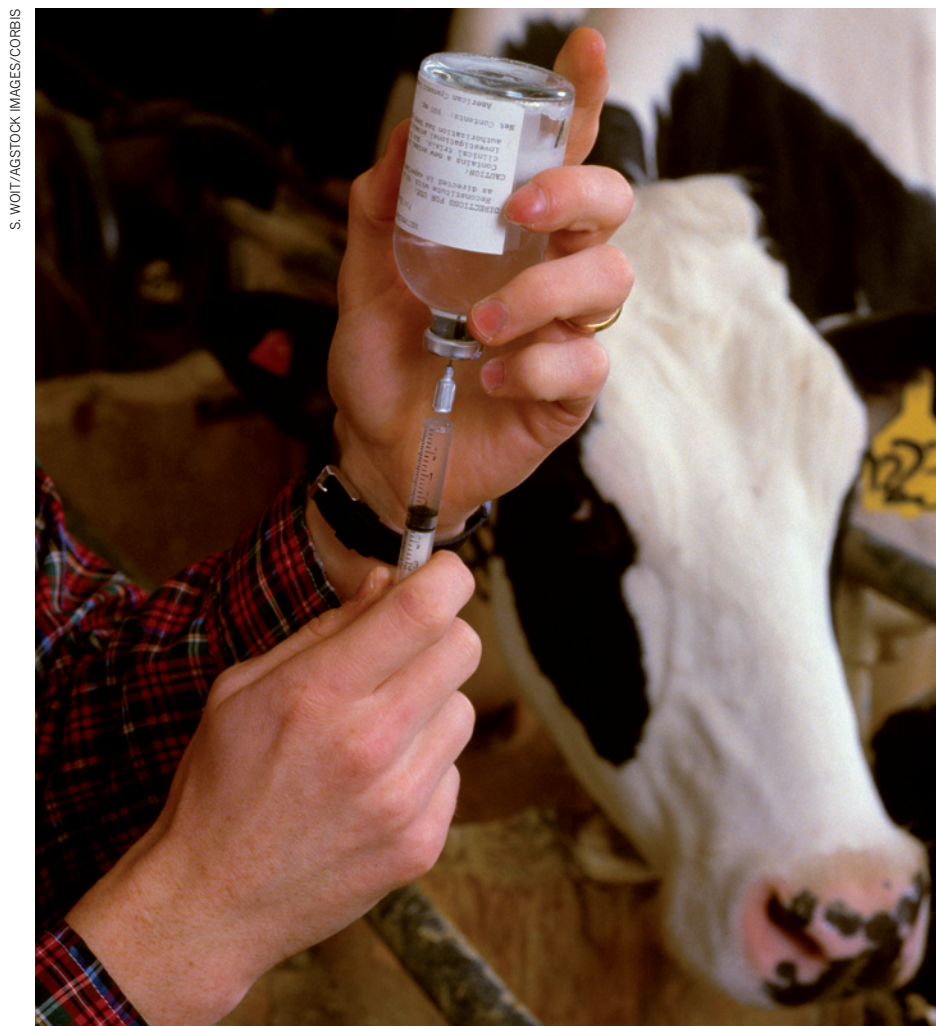
Author Information Coordinates and structure factors for the structures presented here have been deposited in the Protein Data Bank under accession 4DOJ and 4AIN. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.Z. (christine.ziegler@biophys.mpg.de).

CAREERS

TURNING POINT Evolutionary pharmacologist reaches out online **p.133**

FUNDING Crowd-funding gains traction amongst researchers **p.133**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



S. WOIT/AGSTOCK IMAGES/CORBIS

HIGHER EDUCATION

Beyond the farm

Veterinary expertise is an advantage for researchers hoping to stem disease outbreaks and bolster food safety.

BY AMY MAXMEN

In 2002, Newcastle disease spread rapidly through chickens in Los Angeles, California. More than 6,500 poultry died in the effort to control the outbreak. “It took virtually the entire available veterinary staff at the USDA [US Department of Agriculture]

to isolate and eradicate the disease in three months,” says Alan Kelly, dean emeritus of the School of Veterinary Medicine at the University of Pennsylvania in Philadelphia. He fears that if highly infectious foot-and-mouth disease ever reached the United States — it hit the United Kingdom in 2007 and Japan in 2010 — the nation might not have enough

veterinarians to contain it. The trend towards hot summers in some areas could exacerbate viral epidemics, providing favourable conditions for the spread of zoonotic diseases such as West Nile virus and hantavirus.

With the notable exception of clinical practice, shortages haunt pretty much all the sectors in which veterinarians typically look for jobs. Within government, academia and industry, alike, positions for veterinarians with master’s degrees or PhDs remain vacant, according to a report by the US National Research Council (NRC) published at the end of May and written by a committee chaired by Kelly. The committee was formed after members of the US Congress expressed concerns about whether the nation had enough personnel to monitor zoonotic disease outbreaks, regulate food safety and evaluate medicines. “The Department of Labor predicts a lot of opportunities for veterinarians, and I think many of those opportunities will open up in the near future,” says Michael Gilsdorf, executive vice-president of the National Association of Federal Veterinarians, based in Washington DC.

Veterinarians in research-related posts say that many people are not aware of the versatility of a veterinary career. “People have a very narrow view of what veterinarians do,” says Bonnie Buntain, a public-health expert at the University of Calgary in Canada who started out as a horse veterinarian. “Who would have thought that a horse vet from Hawaii would be guiding national regulations on food safety and humane animal treatment in Washington DC at the USDA, and then be offered a tenured professor position?” Those who like to apply analytical skills to real-world problems in research laboratories, the environment, the food supply and beyond will find ample opportunities — but they may have to invest in some extra schooling.

DETECTIVE WORK

Research veterinarians have much to offer. Federal and state governments need them to identify animals with disease; enforce humane slaughter regulations; track wildlife; monitor *Escherichia coli* and *Salmonella* in food; detect the effects of toxic compounds in the ecosystem; inspect dog-breeding facilities and more. Most of the veterinarians involved in government work are employed by the USDA. More-specialized positions are available at the National Institutes of Health (NIH), the Food and Drug Administration (FDA) and the Centers for Disease Control and Prevention ►

► (CDC). At the NIH in Bethesda, Maryland, veterinarians preside over animal facilities, or study cancer and other diseases in animals. At the FDA, based in Silver Spring, Maryland, they investigate food safety, and help to regulate the genetically modified (GM) animals that are slowly making their way through the FDA's regulatory pipeline. And at the CDC in Atlanta, Georgia, they monitor zoonotic disease outbreaks.

Internationally, industry positions for people with a doctorate in veterinary medicine (DVM) and a master's or PhD in toxicology, pharmacokinetics or another field are becoming ever more lucrative. Jobs at pharmaceutical and biotechnology companies pay base salaries ranging from US\$85,000 to \$150,000, depending on the level of seniority. Mary McConnel, a strategic initiatives director at Pfizer, says that her combination of a DVM and a master's in business seemed to be in demand. As soon as she sent her CV to Pfizer, "they were after me like a bat out of hell", she says. McConnel enjoys her job as a consultant for veterinary businesses, and the generous salary that goes with it.

In general, those with a DVM and a science degree have a variety of options in pharmaceutical and biotechnology companies. They might conduct experiments on the efficacy and safety of drugs in animals before human trials begin, or they might help to develop treatments for animals. At animal-supply companies, they often develop and care for GM laboratory animals. At diagnostic laboratory companies, they develop diagnostic tests for both pets and laboratory animals.

SHORTFALL SOLUTIONS

For all their usefulness, however, research veterinarians have been on the decline — especially in government. According to a 2009 report by the US Government Accountability Office, the number of veterinarians working for the federal government had fallen by 40% since 1990. And one-third of veterinarians employed by the USDA, the FDA and the US Army were due to retire in 2011.

Meanwhile, veterinary schools in the United States and the United Kingdom say that they struggle to find candidates that have both clinical and research experience. According to the NRC report, roughly 11% of the veterinarians with faculty positions will be retiring by 2016. Plus, student enrolment at veterinary schools is increasing, which is pushing up the demand for teachers. At the University of Glasgow, UK, Nicholas Jonsson says that the university is now having to hire veterinarians who do not have PhDs. "In the past that would have been unthinkable," he says, "but we desperately need faculty for teaching and clinical posts."

Although most of Jonsson's students pursue careers in pet medicine, he encourages them to consider a career that goes beyond the day-to-day routine of clinical practice.

Jonsson started his career as a farm-animal

veterinarian in rural Australia. For him, it was a less-than-ideal job, he says. "You're on call every other weekend, you drive a lot, you're kicked and beaten and stomped." He then returned to university for a PhD, and later moved to Glasgow, where he researches the evolution of drug resistance in parasites.

US agencies have made some effort to entice new talent. In 2003, the NIH's National Cancer Institute launched a programme for graduate-level biomedical education in partnership with DVM programmes at veterinary colleges across the country. The seven alumni who have completed the NIH Comparative Biomedical Scientist Training programme have gone on to work as postdocs, tenure-track assistant professors, NIH staff scientists and industry pathologists. And in 2008, the CDC introduced a two-year research-focused residency programme with the aim of addressing a shortage of veterinarians to monitor disease outbreaks.

DUAL-DEGREE ADVANTAGE

The path to many research-veterinarian opportunities entails a dual degree, usually with a focus on both veterinary medicine and a basic science such as toxicology, genetics, epidemiology or parasitology. Of the 28 veterinary schools in the United States, 13 offer joint DVM-PhD programmes.

The price of veterinary school alone — around \$66,000 per year for a four-year degree in the United States, and under half that in the United Kingdom for UK citizens — presents an obstacle. After accumulating \$140,000 in debt, Kelly says, young veterinarians tend to want to start earning money rather than enrol in graduate programmes that offer, at best, modest stipends. "It's a desperate situation that has to change," Kelly says. "Salaries need to increase and the cost of veterinary education has to

decrease." The NRC report recommends that the government, economists, industry, deans and veterinary organizations create strategies to reduce the costs of veterinary schools — such as sharing facilities and starting online courses.

Grants from the NIH, called T32 institutional training awards, provide salaries for pre- and postdoctoral DVMs pursuing biomedical research. Merck, Pfizer and other large drug companies offer fellowships to support DVMs who want to return to a research career.

Despite the high price tag of veterinary school, Claude Nagamine says that pursuing the DVM was his best career move — he relishes not only the frequent interaction with animals but also the job security. He decided to pursue a DVM after failing to earn tenure as an assistant professor in the cell-biology department at Vanderbilt University in Nashville, Tennessee.

As he pondered his next move, he says, he realized he loved working with animals. He had spoken with veterinarians working in the animal facility and was intrigued.

Immediately after veterinary school, he began a research residency at an animal-care facility at the Massachusetts Institute of Technology in Cambridge, and ended up at Stanford University in California, where he not only conducts his own research, but also helps scientist colleagues to navigate the regulations and paperwork involved in animal experimentation. "If you're driven to do publishable research, there are lots of jobs out there," he says.

Research veterinarian Anne Fairbrother says that her work combines research with investigative pursuits. Fairbrother, who holds a DVM and a PhD in wildlife disease, worked at the US Environmental Protection Agency for 13 years, monitoring the risks posed by GM plants on wildlife. This often involves checking the concentration of chemicals in soil, plants and water, and then comparing them with concentrations in animal serum and tissue.

Or, if animals start to die in abnormal numbers, she might be asked to search for clues related to parasite or viral infections, compounds such as petroleum that might have leaked from gas stations and signs of industrial pollutants. "If you go to a place where there is contamination, you know what questions to ask to find out the history of the area and which signs to look for, because sometimes chemicals are causing problems but it could be something else," she says. "That diagnostic approach is something that you learn in veterinary school."

But a love for animals is probably the main requirement for researchers looking to multiply their career options with a DVM degree. At a time when degrees and certifications rarely confer job guarantees, the veterinary research option stands apart. ■



Nicholas Jonsson encourages students to go beyond the usual career choice of pet medicine.

Amy Maxmen is a locum biology correspondent at Nature.

TURNING POINT

Ethan Perlstein

Ethan Perlstein has spent five years creating a sub-field of research that he calls evolutionary pharmacology. As his fellowship at Princeton University's Lewis-Sigler Institute for Integrative Genomics in New Jersey comes to an end and he searches for his next academic post, Perlstein is maintaining an innovatively designed website adorned with modules and discussion threads to help to communicate his thoughts on science.

When did you first start fostering communication among scientists?

I was an intern at a small biotech company before my final year of high school. As part of that, I would read immunology articles, formulate questions and start a correspondence with the author. One of these authors was Ronald Germain, an immunologist at the US National Institutes of Health. He must have been struck by the idea of a kid reading papers; he offered me another internship, in his lab in Bethesda, Maryland, where I worked for the summer before going to Columbia University in New York to study sociology.

How did you come to champion evolutionary pharmacology?

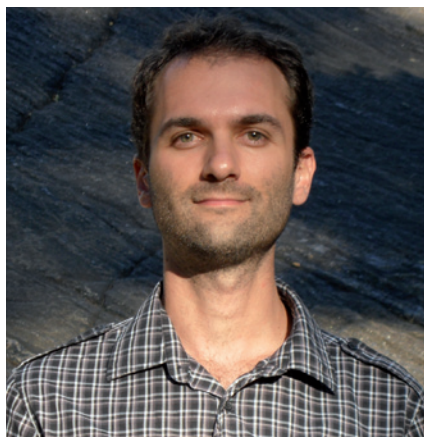
After several rotations in different labs as a cell-biology graduate student at Harvard University in Cambridge, Massachusetts, I realized that I wanted to work with small molecules relevant to human diseases. I also wanted to use a simple model system, such as yeast, so that I could do a lot of experiments quickly. I noticed that several small molecules that affect yeast growth are also psychiatric drugs, and I started studying the connection. There is a large evolutionary distance between yeast and humans, but these drugs affect ancient processes that we share.

How is your job search going?

Like many of my colleagues, I have been battered by the job market. I received zero interviews out of 18 applications this year, despite having a five-year independent position with a US\$1-million budget on my CV. The fellowship has been great, but it is not a normal postdoc, so I am not sure that the wider community of scientists knows what to make of it. I am a research cross-pollinator, and don't have one well-known area of expertise. I wonder if it may be harder for me to break out.

Has your decision to invest time, money and energy in your website paid off?

I think so. Most academics use a lab website



to list publications; essentially it becomes static, a version of their CV. I wanted to do something new and cool that would help to communicate science. I am not a programmer, so I spent several thousand dollars of my own money on hiring a professional design team to create something interactive. It includes my tweets, blog posts and research summaries — replete with pop-culture references — in a series of modules that encourage viewers to add their own comments. It seems to work: one private-sector researcher who checked out my website contacted me about mutually beneficial research opportunities.

One post on your website breaks down your academic lab budget. Why share this?

My fellowship finishes at the end of the year, and I am interested in crowd-funding a project on how amphetamines such as crystal meth work. I am asking for roughly \$25,000, and I thought that I should give potential funders evidence that I am responsible with money. I see this as part of the same movement as a group of scientists who are posting their grant proposals — whether they are successful or not. I am excited about experimenting with the way we do science.

What has been your career turning point?

Without a doubt, joining Twitter in 2011, when I started offering my thoughts about changing the way science is done. I found a community of people passionate about rethinking scholarly publishing and funding. I had hoped for a way to scale up the e-mail cold-calling that I had done at high school. Twitter was a way to connect with like-minded people and keep a conversation going 24/7. ■

INTERVIEW BY VIRGINIA GEWIN

FUNDING

People power

Researchers are starting to turn to crowd-funding as a way to support their work, says Simon Vincent, head of personal awards and training at Cancer Research UK in London. At the 20 September Naturejobs Career Expo in London, Vincent said that in a lean funding environment, seeking donors can be easier than navigating the grant-application process. "There's no peer review or middlemen," he told conference attendees. "If you have a good idea and can convince enough people, you get the money." But Vincent warned that crowd-funding — seeking funds through the online community — also has downsides. Traditional research-grant peer review provides quality control, a reality check and a way to hone and refine an idea, and the interaction with the funder can provide links to large, established networks in the scientific community. Crowd-funding, even with established sites such as petridish.org, requires a lot of time and public interaction, Vincent said. Scientists often have to make a video about their research project and must stay in regular contact with donors, who can number in the hundreds.

ETHICS

Relationship advice

A university conflict-of-interest committee should review contracts between academic scientists and industry sponsors that are worth US\$5,000 or more, concludes a draft report entitled Recommended Principles & Practices to Guide Academy-Industry Relationships. Researchers should never ghostwrite research papers and should retain oversight of intellectual property and a stake in the proceeds from patents, according to the proposal. The report offers 56 guidelines for maintaining academic freedom and upholding ethical conduct in partnerships and collaborations between academics and industrial sponsors. Issued on 18 September, it was written by the American Association of University Professors in Washington DC in response to the increasing number and complexity of such partnerships, says co-author Cary Nelson, a past president of the association. "The corrupting power of money has become much more clear," he says, noting that issues such as sponsors suppressing data from studies and persuading eminent researchers to add their names to papers they did not write seem to be on the rise.

MAN'S BEST FRIEND

Animal instincts.

BY GRACE TANG

“Come on girl, you can do it.” I gently coaxed Dr Gleitman’s latest subject as she shook off the last of the drugs and struggled to lift herself from the bed. At three years of age, Callie was Dr Gleitman’s youngest subject to date. Her dark brown eyes glinted as she gradually blinked them open, adjusting to the harsh fluorescent lights of the post-op recovery room.

The matte black chassis of the neural implant peeked out from a small bald patch between the fine gold strands on her head. Even though I was not the one who put it on her, as Dr Gleitman’s research assistant, I felt a twinge of guilt that I was subjecting her to this painful procedure — without her consent, no less, as she was unable to give it before the operation. But then again, God did not obtain consent when he created Man from the clay.

And now Man was imparting God’s greatest gifts to his best friend. I’d spent a good portion of my working life in this lab, and even though I’d seen countless animals pass through these halls, I’d grown fond of Callie since we’d got her from the local pound, a day before she was scheduled to be put down. She was up on her feet now. She sniffed cautiously at me.

I wondered what vocation she would be assigned to as I leaned in close to her and let her lick my face. Because our funding came from the military, primates were usually used in jungle warfare. Cats, with their excellent night vision and stealth, were used in reconnaissance. Dogs usually went to the army or police, for more traditional roles as sniffers or for search-and-rescue. With her gentle demeanour, Callie probably would not be an attack dog. I heard that augmented animals were being used for therapeutic procedures now. Maybe she would be trained to be a seeing-eye dog, or used in the treatment of post-traumatic stress disorder?

“Your name is Callie. Can you say Callie?”

Of course I was not expecting her to *speak*, in the strictest sense of the word. Her lips were not shaped for speech, and unlike the apes, her limbs were not shaped for signing. Her ears perked up, and she tilted her head in the most endearing way.

“Cal... lie...?”

The single red LED

GO NATURE.COM
Follow Futures on
Facebook at:
go.nature.com/mtoodm

on the speech synthesizer blinked, indicating that the implant had successfully extracted, from the spatial and temporal firing patterns across hundreds of thousands of Callie’s neurons, her thoughts, emotions and intentions, and further transduced those signals to spoken words, complete with affective tone, closely mimicking human speech. Dr Gleitman had added yet another success to his list.



I darted over to the adjoining office where Dr Gleitman was asleep in his reclining chair, feeling Callie’s inquisitive eyes on me as I left the room. Normally, I would not disturb him, but I thought he might want to know that his latest subject was awake and talking. I nudged his hand. He jumped slightly, disoriented for a second.

“What is it, Moe?”

“Callie’s unit is already functional, Sir! She managed to say her name when I asked her to.”

The poorly masked irritation at having his nap interrupted melted away into a satisfied smile.

“So quickly? That’s incredible. She must be the most talented canine we’ve had yet. Usually it takes at least a full day for speech comprehension to begin, let alone production of the first word.”

He rose from his chair and walked to where Callie was busy pawing at her

Elizabethan collar. “I wonder if it’s because her youth makes her brain more malleable and lends itself better to the implant. Or perhaps it’s because she’s female. Moe, can you make a note of this?”

I went to the corner of the lab where the video recorder was and made notes for the day, watching Dr Gleitman interact with Callie as I narrated my observations.

“Hi Callie, do you understand me? Do you know where you are?” Dr Gleitman was already trying to extract complete sentences from her. That usually took weeks. Clearly he expected more from her compared with his previous subjects.

When I was done, I eagerly returned to Callie’s bedside. I worked up the courage to ask Dr Gleitman the question that had been burning inside me ever since we brought her home.

“Do you think you will keep her?”

He looked shocked — perhaps I had spoken out of place. Surely he would say no, and Callie would be deployed to some faraway base, where I’d never see her again. But his face softened.

“I’ve actually thought about it. Because of her exceptional performance, I could easily argue for her to be kept here for observation and testing. Not to mention, I’m developing a soft spot for you Golden Retrievers.”

I could barely keep still in my excitement, until I remembered what Dr Gleitman had said about behaving more like a human and not like a stray pup if I wanted to keep my job in the lab. But it was hard keeping my tail still when the news was making Callie’s tail wag so fast that she was sending strands of fur flying. Dr Gleitman was grinning widely at the sight. In time he would teach her to use language to convey her thoughts instead of these primal displays of emotion, as he had done with me.

“Well I’m pleased both of you are happy with this arrangement. You’ll have to do the paperwork though, Moe.”

“Gladly!” I would get started on that later, but for now I thought of what to say to welcome the latest addition to our home. ■

Grace Tang is a graduate student in psychology at Stanford University. One of her greatest wishes is to have a talking animal sidekick.

JACEY